VISTAS IN ASTRONOMY

Volume 3

" . . . and I need scarcely remind you tnat in the detective story which is science, clues add up non-linearly: two and two can add up to more than four . . . "

(J. L. PAWSEY in *Vistas in Radio-Astronomy*, p. 2, Sydney, 1957.)

Special Supplement No. 7 to the Journal of Atmospheric and Terrestrial Physics

VISTAS IN ASTRONOMY

EDITED BY

ARTHUR BEER

Volume 3

PART I

DYNAMICS — GEOPHYSICS
INSTRUMENTS — SOLAR SYSTEM

PART II

STELLAR ASTRONOMY — STELLAR EVOLUTION
PHOTOMETRY — GALAXIES
SPECTROSCOPY — COSMOLOGY

PERGAMON PRESS
LONDON · OXFORD · NEW YORK · PARIS
1960

PERGAMON PRESS, LTD.
4 & 5 Fitzroy Square, London W.1
Headington Hill Hall, Oxford

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue, N.W., Washington 5, D.C.
P.O. Box 47715, Los Angeles, California

PERGAMON PRESS S.A.E.L.
*24 Rue des Écoles, Paris V^e^*

PERGAMON PRESS G.m.b.H.
Am Salzhaus 4, Frankfurt

First Published 1960

Copyright
©
1960
PERGAMON PRESS LTD.

VO
3
19

Library of Congress Card No. 56-4213

Printed in Great Britain by *J. W. Arrowsmith Ltd.*, Bristol

INTRODUCTION

THE two volumes of *Vistas in Astronomy*, originally conceived out of a desire to do honour to Professor STRATTON, have received so warm a welcome that we have been encouraged to continue these research surveys as a permanent series.

New "vistas" have opened on the astronomical horizon, and old ones have changed in colour and perspective.

We considered it essential to maintain the international character of the original volumes. The present volume, and two others now in preparation, contain contributions from America, Belgium, Canada, China, Czechoslovakia, France, Germany, Great Britain, India, Italy, Mexico, the Netherlands, Portugal, Russia, and Sweden. We are especially pleased that our Russian colleagues are making available a regular series of reports on the progress of Soviet Astronomy.

Arrangements have been made to ensure that at the time of publication each new volume of *Vistas* will present a thoroughly up-to-date picture.

The Observatories,
Cambridge.

ARTHUR BEER.

CONTENTS OF VOLUME 3

PART I

	page
The Dynamics of Stars in the Neighbourhood of the Sun *by* R. v. d. R. Woolley, F.R.S., Astronomer Royal (Herstmonceux)	3
The Rediscovery of Neptune *by* R. A. Lyttleton, F.R.S. (Cambridge, England)	25
The Determination of Einstein's Light-Deflection in the Gravitational Field of the Sun *by* H. von Klüber (Cambridge, England)	47
The Clock Paradox *by* W. Cochran (Cambridge, England)	78
On the Clock Paradox *by* V. Fock (Leningrad, U.S.S.R.)	88
On the Design of Large Telescopes *by* G. M. Sisson (Newcastle-upon-Tyne)	92
The Electronic Computer as an Astronomical Instrument *by* Marshal A. Wrubel (Bloomington, Indiana)	107
Techniques in Auroral Investigations *by* A. H. Jarrett (Belfast)	117
Note on Optical Observations of the Russian Satellites *by* D. W. Beggs (Cambridge, England)	126
The Two Bright Comets of 1957:	
I Arend-Roland (1956 h) and Mrkos (1957 d) *by* J. G. Porter (Herstmonceux)	128
II Cometary Physics and the Comets Arend–Roland and Mrkos *by* N. Richter (Sonneberg, Germany)	132
III The Anomalous Tail of Comet Arend–Roland *by* Gunnar Larsson-Leander (Saltsjöbaden, Sweden)	137
IV Filter-Photography of the 1957 Comets *by* Neville J. Woolf (Manchester)	143

PART II

The First 70,000 *by* John B. Irwin (Bloomington, Indiana)	147
Multiple Colour Photometry *by* Gerald E. Kron (Mount Hamilton, Cal., U.S.A.)	171
Experiences in Photometry with the Cambridge Schmidt Telescope *by* A. N. Argue (Cambridge, England)	184
The Excitation of Emission Lines in the Spectra of Early-Type Stars *by* B. E. J. Pagel (Herstmonceux)	203
Colour, Luminosity and Evolution of the Stars *by* Olin J. Eggen (Herstmonceux)	258
The Visual Milky Way *by* Sergei Gaposchkin (Cambridge, Mass., U.S.A.)	289
Star Formation and the Galactic Magnetic Field *by* L. Mestel (Cambridge, England)	296
Observational Aspects of Cosmology *by* D. W. Sciama (London)	311
Contents of Volumes 1 and 2	329
Name Index	335
Subject Index	341

NAME INDEX

Abell, G. O., 313, 324, 326
Adams, A. N., 154, 170
Adams, J. C., 25, 26, 27, 29, 30, 36, 37, 38, 39, 44, 45, 46
Ahnlund, K., 325, 326
Ahrens, L. H., 325, 326
Airy, G. B., 35
Alexander, J. B., 15, 17, 19, 24
Alfvén, H., 305
Allen, C. W., 228, 255
Aller, L. H., 205, 211, 212, 221, 222, 226, 231, 232, 236, 237, 242, 248, 255, 256, 311, 325, 326
Alpher, R. A., 324, 325, 327
Ambarzumian, V. A., 246, 247, 248, 255
Anderson, A., 49, 65, 75
Arend, S., 128
Argue, A. N., 144, 184
Armstrong, E. B., 123, 124, 129, 132
Arp, H. C., 183, 285
Aten, A. H. W., 325, 327
Atkinson, R. d'E., 104

Baade, W., 23, 147, 313, 314, 315, 319, 327
Babcock, H. D., 121, 124
Babcock, H. W., 164, 167, 169, 227, 255
Baker, J. G., 205, 211, 226, 230, 231, 232, 233, 234, 236, 237, 242, 248, 249, 255, 256
Baldwin, J. E., 328
Balmer, J. J., 209–254
Bappu, M. K. V., 285
Barney, I., 157, 169, 170
Bauer, L. A., 49, 65, 75
Baum, W., 285, 286, 313, 314, 319, 321, 327
Beals, C. S., 203, 214, 217, 218, 219, 220, 221, 222, 223, 224, 225, 227, 229, 251, 254, 255
Becker, W., 173, 174, 176, 180, 183, 185, 188, 202
Beer, A., 44, 76, 106, 141, 142, 169, 170, 202, 255, 256, 257
Beggs, D. W., 126, 127
Bennett, R., 157, 170
Bernouilli, J., 302
Bessel, F. W., 131, 132, 139
Bethe, H. A., 324, 326
Beyer, M., 144
Beynon, W. G., 256
Biber, C., 185, 202
Bidelman, W. P., 178, 183
Biela, W. von, 141, 142
Biermann, L., 133
Bigay, J. H., 143, 144
Blaauw, A., 148, 152, 169, 286
Blackwell, D. E., 143, 144
Blanco, V., 286
Bloch, M., 227, 256
Bode, J. E., 25, 29, 35, 36, 37, 38
Bok, B. J., 23, 24
Bolton, J. G., 316, 327
Boltzmann, L., 231, 238
Bondi, H., 286 299, 301, 302, 303, 304, 305, 310, 311, 312, 313, 322, 326, 327
Bonnor, W. B., 323, 327
Bonsack, W. K., 286
Born, M., 80, 86
Bottlinger, K. F., 49, 50, 75
Bowen, I. S., 164, 169, 212, 215, 216, 217, 220, 255
Boyartchuk, A. A., 243, 245, 253, 255
Bracewell, R. N., 327
Bredichin, T., 130, 132, 141
Bremiker, C., 44
Brown, A., 286
Brown, E. W., 45, 46
Brown, G. M., 256
Brownlee, R., 286
Brunn, A. von, 51, 54, 55, 56, 58, 66, 67, 68, 69, 70, 71, 75, 76
Bucerius, H., 50, 75
Builder, G., 80, 82, 86
Bullough, K., 124
Burbidge, E. M., 204, 209, 210, 219, 226, 227, 228, 229, 230, 233, 243, 251, 255, 325, 327
Burbidge, G. R., 204, 209, 210, 219, 226, 227, 228, 229, 230, 233, 243, 251, 255, 312, 325, 327
Burr, E. J., 313, 314, 327
Byard, P. L., 124

Cameron, A. G. W., 312, 325, 327
Campbell, W. W., 48, 49, 58, 65, 67, 68, 75
Candy, M. P., 131, 132
Challis, J., 25, 35
Chalonge, D., 160, 169
Chamberlain, J. W., 235, 236, 255
Chandler, S. C., 171
Chandrasekhar, S., 111, 204, 207, 209, 218, 222, 247, 255, 260, 275, 286
Chant, C. A., 58, 65, 75
Chapman, S., 117, 123, 124, 125
Cherbak, P., 131
Chuvajev, K. K., 287
Cillié, G., 219, 220, 226, 236, 248, 251, 256
Clegg, J. A., 124, 125
Clerke, A., 259
Cochran, W., 78, 80, 83, 86
Code, A. D., 148, 170, 322, 328
Cojan, J., 122, 125
Comas y Solá, J., 67, 75
Cook, C. W., 325, 327
Couder, A., 95, 184, 202

Courvoisier, L., 49, 75
Coutrez, R. A. J., 134
Cowling, T. G., 300
Cox, A. N., 286
Cox, H. W., 94
Crawford, F. S., 80, 86
Crawford, J. A., 286
Crommelin, A. C. D., 49, 64, 65, 75, 76
Cullwick, E. G., 80, 86
Currie, B. W., 124, 125
Curtis, H. D., 48, 75

Dalgarno, A., 124
Danjon, A., 50, 57, 59, 65, 66, 67, 69, 72, 75, 92, 106
Darsenius, G., 138, 139, 142
Darwin, C., 80, 86
Davidson, C. R., 58, 64, 65, 75
de Jager, C., 208, 256
de Sitter, W., 313
de Vaucouleurs, A., 172, 183
de Vaucouleurs, G., 322, 324, 328
Delhaye, J., 148, 169
Deutsch, A., 286
Dewhirst, D. W., 185, 195, 202
Dieter, N. H., 317, 321, 327
Dines, W. H., 49, 65, 75
Dingles, H., 80, 81, 82, 86
Doan, N. H., 143, 144
Dodwell, G. F., 58, 65, 75
Doppler, J. C., 85, 233
Dufay, M., 131, 132, 143, 144
Dunbar, D. N. F., 325, 327
Durbin, R. P., 83, 86
Dyer, E. R., 288
Dyson, F. W., 50, 58, 64, 65, 75

Eddington, A. S., 20, 49, 50, 58, 64, 65 75, 76, 80, 86, 115, 206, 210, 246, 256, 259, 266, 319, 324, 326
Edge, D. O., 312, 315, 316, 327
Eggen, O. J., 20, 22, 24, 175, 183, 258, 296, 288, 313, 314, 319, 327
Einstein, A., 47, 48, 50, 59, 60, 72, 76, 78, 79, 80, 81, 82, 85, 86, 249, 311, 313, 317, 318, 319, 320, 325
Elbert, D., 111
Ellison, M. A., 92, 106
Ellyett, C. D., 124, 125
Elsmore, B., 328
Emden, R., 49, 76
Encke, J. F., 44
Esclangon, E., 76

Farkas, L., 325, 327
Fehrenbach, Ch., 94
Fellgett, P. B., 164, 167, 169, 193, 202
Ferraro, V. C. A., 305
Ferrier, M., 49, 76
Field, G. B., 312, 320, 321, 327, 328
Fischer, H., 134
Fjodorov, E. P., 56, 72, 77
Fock, V., 88
Fogelquist, R., 138, 139, 141, 142
Forbes, G., 156, 169
Forsyth, A. R., 50, 76
Forsyth, P. A., 124, 125
Foucault, L., 100
Fowler, A., 259
Fowler, W. A., 312, 325, 327
Fremlin, J. H., 80, 86
Freundlich, E. F., 48, 50, 51, 54, 55, 56, 58, 64, 65, 66, 67, 68, 69, 70, 71, 72, 76

Gadsden, M., 120, 125
Galileo Galilei, 260
Galle, J. G., 44
Gamow, G., 323, 324, 326, 327
Gaposchkin, S., 222, 248, 256, 289, 293
Gascoigne, S. C. B., 159, 169, 175, 176, 180, 183, 287, 313, 314, 327
Gerasimović, B. P., 218, 256
Gill, D., 156
Giovanelli, R. G., 208, 256
Gleich, G. von., 49, 67, 76
Gleissberg, W., 54, 76
Gliese, W., 15, 18, 23, 24
Gold, T., 312, 313, 322, 327
Goldberg, L., 205, 240, 256
Gollnow, H., 49, 76
Gordon, K., 222, 223, 246, 256
Götz, W., 136, 137
Goudsmit, S., 83
Greaves, W. M. H., 172, 183
Greenstein, J. L., 135, 286, 325, 327
Grubb-Parsons, H., 92, 195
Guggenheimer, K. M., 325, 327
Gyldenkerne, K., 288

Haffner, H., 185, 186, 194, 195, 202
Hagemann, G., 49, 76
Hale, G. E., 258, 260
Hall, J. S., 166, 169, 173, 175, 183
Halley, E., 131, 132, 133
Hallwin, H. D., 50, 76
Harang, L., 120, 121, 124, 125
Harding, K. L., 141, 142
Hare, C., 131
Hargreaves, F. J., 94
Haro, G., 322, 327
Harris, D. L., 112, 176, 183, 286, 287
Harteck, P., 325, 327
Haségawa, I., 128, 132
Haselgrove, C. B., 286, 319, 322, 327
Haser, L., 133
Hatcher, R. D., 218, 255
Havens, W. W., 83, 86
Hazard, C., 317, 327
Heard, J. F., 160, 169
Hebb, M. H., 205, 256
Heckmann, O., 150, 156, 169, 311, 326
Heeschen, D. S., 312, 317, 327
Heitler, W., 325, 327

Heller, L., 325, 327
Helmholtz, H. von, 258, 275, 299, 306, 310
Henyey, L. G., 282, 286
Hepperger, J., 50, 76
Herbig, G. H., 282, 286
Herman, R. C., 325, 327
Hermann, H., 50, 76
Herschel, W., 260
Hertzsprung, E., 262, 266, 269, 276
Herzog, E., 324, 327
Hewish, A., 316, 328
Hiltner, W. A., 158, 166, 167, 169, 175, 177, 183, 223, 224, 256, 287
Hipparchus, 171
Ho, Ping-Yü, 141, 142
Hogg, A. R., 286
Hopmann, J., 50, 65, 67, 76
Hoshida, G., 72, 77
Hoyle, F., 260, 275, 286, 312, 313, 318, 319, 320, 322, 324, 325, 327
Huang, Su-Shu, 286
Hubble, E. P., 315, 319, 321, 326
Humason, M. L., 313, 314, 327
Hunaerts, J., 134
Hunten, D. M., 123, 125
Hynek, J. A., 256

Inglis, D. R., 228
Irwin, J. B., 147, 159, 163, 167, 169
Isaak, G. R., 83, 86
Ives, H. E., 80, 86

Jacchia, L. G., 126
Jackson, J., 69, 76
Jaegermann, R., 139, 141, 142
Jánossy, L., 83, 86
Jarrett, A. H., 117, 121, 124, 125
Jeans, J. H., 259, 324, 326
Jeffreys, H. J., 49, 65, 76
Jenkins, L. F., 162, 169
Johnson, H. L., 112, 158, 159, 160, 167, 169, 174, 175, 176, 183, 188, 202, 260, 287
Jonckheere, R., 49, 76
Junkes, J., 94
Just, R., 323, 327

Kaiser, T. R., 124, 125
Kapteyn, J. C., 156
Karandikar, R. V., 123, 125
Keeping, E. S., 223, 256
Kelvin, Lord (Sir William Thomson) 299, 306, 310
Kemmer, N., 88
Kepler, J., 204, 309
Kienle, H., 130, 181, 182, 183
King-Hele, D. G., 126
Klüber, H. von, 47, 51, 56, 57, 58, 66, 67, 68, 69, 70, 71, 73, 75, 76
Koeckelenbergh, A., 134
Kogure, T., 221, 256
König, A., 52, 57, 76
Kopal, Z., 222, 223, 256
Kopff, A., 50, 76
Kosirev, N. A., 204, 256
Kourganoff, V., 214, 256
Kraus, J. D., 134
Kron, G. E., 159, 169, 170, 171, 174, 175, 176, 178, 179, 180, 183, 222, 223, 246, 256, 260, 261, 286, 287
Kuiper, G. P., 267, 269, 275
Kukarkin, B. V., 287
Kuragano, S., 131
Kushwara, R. S., 287

Lallemand, A., 164, 176
Land, G., 155, 170
Landmark, B., 124, 125
Langevin, P., 80, 86
Larsson-Leander, G., 129, 132, 137, 142, 143, 144
Laue, M. von, 48, 76, 80, 86
Lauritsen, C. C., 325, 327
Lauritsen, T., 325, 327
Layzer, D., 312, 323, 324, 327
Lebedinsky, A. I., 123, 125
Ledermann, W., 54, 76
Lemaître, G., 319, 323, 327
Lentz, J., 157, 170
LeLevier, R., 286
Levée, R., 286
Le Verrier, U. J. J., 25, 26, 27, 29, 30, 36, 37, 38, 39, 44, 45, 46
Lewis, J. M., 49, 76
Lilley, A. E., 317, 320, 327
Limber, D. N., 324, 327
Lindblad, B., 8
Lindblad, P. O., 142
Lindemann, A. F., 48, 76
Lindemann, F. A., 48, 49, 76
Linfoot, E. H., 184, 202
Liouville, J., 4
Little, C. G., 316, 328
Littlewood, J. E., 26, 27
Loar, H. H., 83, 86
Lockyer, N., 258
Lodén, K., 142
Lodén, L. O., 142
Lorentz, H. A., 78, 80, 87
Lovell, A. C. B., 124, 125, 312
Ludendorff, H., 69, 77
Lüst, R., 309, 310
Lyman, Th., 205–252
Lyons, H., 85, 87
Lyons, U. S., 154, 170
Lyttleton, R. A., 25

Markowitz, W., 155, 170
Martel, M., 143, 144
Masevich, A. G., 287
Mathis, J. S., 286
Mattig, W., 50, 56, 69, 77
Matukuma, T., 58, 72, 77
Mayall, N. U., 175, 176, 179, 313, 314, 322, 328

McClain, E. F., 317, 320, 328
McCrea, W. H., 80, 85, 87, 287, 311, 312, 313, 320, 321, 327, 328
McCuskey, S. W., 20, 24
McLaughlin, D. B., 204, 214, 254, 256
McNamara, A. G., 124, 125
McVittie, G. C., 311, 326
Mees, C. E. K., 191, 202
Meinel, A. B., 122, 125
Melbourne, W. G., 286
Meltzer, A., 287, 325, 328
Menzel, D. H., 205, 211, 226, 230, 231, 232, 233, 234, 236, 237, 242, 248, 249, 255, 256
Merrill, P. W., 227, 256, 325, 328
Mestel, L., 287, 296, 306, 310, 324, 328
Mikesell, A. H., 166, 169
Mikhail, F. I., 86, 87
Mikhailov, A. A., 50, 56, 57, 58, 59, 60, 64, 65, 66, 69, 71, 72, 74, 77
Mills, B. Y., 312, 315, 316, 328
Milne, E. A., 115, 206, 245, 256
Minkowski, R., 315, 317, 328
Minnaert, M. G. J., 325, 328
Mitchell, R. I., 287
Miyamoto, S., 205, 221, 247, 256
Moller, C., 80, 81, 87
Morgan, H. R., 152, 154, 155, 169, 170
Morgan, W. W., 148, 158, 160, 167, 169, 170, 174, 175, 176, 183, 188, 202, 204, 256, 286, 287, 314, 322, 328
Mrkos, A., 131
Muller, C. A., 20, 24
Müller, H., 134
Münch, G., 223, 224, 256, 287, 312, 322, 328
Munitz, M. K., 311

Neckel, H., 130
Needham, J., 141, 142
Nekrasova, S. V., 287
Neubauer, F. J., 221, 256
Neugebauer, G., 286
Newall, H. F., 49, 65, 77
Newburn, R. L., 286
Newton, I., 48
Neyman, J., 312, 323, 324, 328
Nicholson, T. D., 69, 77
Nikonov, V. B., 287

Ogorodnikov, K. F., 23, 24
Oke, J. B., 217, 256
Olbers, W., 141, 142
Olson, E. C., 112
Olsen, K. H., 286
Oort, J. H., 5, 19, 20, 24, 272
Öpik, E., 171, 173, 183
Osterbrock, D. E., 131, 132
Ovenden, M. W., 50, 77

Pagel, B. E. J., 22, 203, 219, 228, 230, 256, 312
Pannekoek, A., 211, 228, 237, 256, 290
Papapetrou, A., 50, 74, 77
Parenago, P. P., 287
Paschen, F., 215–252
Paton, J., 117, 124, 125
Pauli, R. T., 325, 328
Pawsey, J. L., I, 317, 328
Perrine, C. D., 48, 50, 77
Peters, W. J., 49, 65, 75
Petrie, R. M., 160, 170
Petrie, W., 124, 125
Pickering, E. C., 251
Pirani, F. A. E., 312
Planck, M., 182
Pokrowski, G. I., 325, 328
Polosuhina, N. S., 287
Poor, Ch. L., 50, 67, 77
Popov, N. A., 56, 72, 77
Porter, J. C., 67, 77
Porter, J. G., 128
Powell, C. F., 325, 327
Priester, W., 134
Proisy, P., 49, 77
Purkinje, J., 172
Purcell, E. M., 321, 328

Rachkovsky, D. N., 287
Rayleigh, Lord (J. W. Strutt), 310
Reddish, V. C., 287
Redman, R. O., 185, 202
Reiz, A., 155
Richardson, L. F., 49, 65, 75
Richter, N., 132
Ring, J., 121, 125
Roach, F. E., 204, 215, 229, 256
Robertson, H. P., 312, 313, 315, 318, 319, 323, 328
Roemer, E., 128, 132
Roland, G., 128
Roman, N. G., 228, 251, 256, 260, 286, 287
Rösch, J., 92, 106
Ross, F. E., 49, 77, 194, 202
Rosseland, S., 208, 210, 216, 242, 256
Rottenberg, J. A., 203, 212, 214, 217, 218, 224, 231, 254, 255, 256
Russell, H. N., 65, 77, 148, 170, 259, 266
Rutherford, E., 282
Ryle, M., 312, 315, 316, 317, 328

Saha, M. N., 231
Salpeter, E. E., 287, 325, 328
Sandage, A. R., 159, 169, 260, 275, 278, 285, 287, 313, 314, 318, 319, 322, 327, 328
Sané, M., 314, 328
Savedoff, M. P., 287
Schalén, C., 142
Scheuer, P. A. G., 315, 316, 327, 328
Schilt, J., 163, 170, 194
Schlesinger, F., 155, 170
Schlüter, A., 309, 310
Schmidt, B., 95
Schmidt, M., 18, 24
Schönberg, M., 286
Schrödinger, E., 311, 326

Schrutka, G., 132
Schueking, E., 314, 328
Schuster, A., 111, 112, 115, 203, 206, 210, 256
Schwarzschild, B., 287, 325, 328
Schwarzschild, K., 16, 24, 84, 87, 111, 112, 115, 269
Schwarzschild, M., 260, 275, 287, 325, 328
Schwesinger, G., 98, 106
Schürer, M., 57, 77
Sciama, D. W., 311, 321, 322, 323, 326, 327, 328
Scott, E. L., 312, 314, 319, 323, 324, 327, 328
Scott, F. P., 155, 170
Searle, L., 287, 325, 328
Seaton, M. J., 235, 256
Secchi, A., 171
Seidel, L. P. von, 93
Selwyn, E. H. W., 191
Serrurier, M., 99
Shakeshaft, J. R., 312, 315, 316, 327, 328
Shane, C. D., 313, 324, 328
Shapley, H., 148, 170
Shapley, M. S., 222, 256
Sharpless, S., 288
Sherman, F., 204, 256
Shortley, G. H., 205, 256
Silberstein, L., 49, 77
Singer, S. F., 84, 85, 87
Sisson, G. M., 92, 106
Slee, O. B., 315, 316, 326, 328
Slocum, F., 49, 65, 77
Smart, M. W., 3, 4, 8, 24
Smith, J. L., 159, 170, 178, 183
Sobolev, V. V., 203, 204, 221, 224, 229, 234, 248, 249, 250, 251, 252, 253, 254, 255, 256
Soldner, J., 47, 77
Sommer, A. H., 159, 170
Sotome, K., 72, 77
Spitzer, L., 287, 306, 310, 315, 324, 326, 328
Stanley, G. J., 316, 327
Stebbins, J., 174, 178, 179, 180, 183, 261, 322, 328
Steinlin, U., 173, 176, 180, 183, 188, 202
Stibbs, D. W. N., 123, 124
Stock, J., 314, 328
Stoffregen, W., 124, 125
Störmer, C., 117, 124, 125
Stoy, R. H., 157, 170, 260, 288
Stoyko, N., 50, 77
Strand, K. A., 162, 170
Stratton, F. J. M., v
Strohmeyer, C. E., 50, 65, 77
Strömberg, G., 15, 18, 19, 24
Strömgren, B., 160, 170, 205, 208, 209, 211, 256, 266, 267, 269, 275, 288
Struve, O., 204, 205, 215, 216, 217, 218, 223, 224, 229, 239, 240, 242, 243, 256, 288
Suess, H. E., 325, 328
Swings, P., 133, 215, 216, 217, 218, 221, 224, 256

Tayler, R. J., 288
Tcheng, Mao-Lin, 227, 256
Teller, E., 228
Ter Haar, D., 309, 310
Thackeray, A. D., 203, 204, 256
Thiessen, G., 144
Thomas, R. N., 224, 256
Thomson, J. H., 328
Thomson, J. J., 205, 223
Thomson, J. V., 94
Tift, W. G., 286
Tikhov, G. A., 50, 77, 172, 173, 183
Tisserand, F., 37
Tolman, R. C., 80, 87
Tonsberg, E., 122, 125
Trumpler, R., 48, 49, 50, 58, 65, 66, 67, 68, 69, 75, 77, 264, 267

Underhill, A. B., 203, 206, 207, 210, 212, 216, 218, 221, 256
Unsöld, A., 241, 257
Urey, H. C., 325, 328

van Biesbroeck, G., 56, 58, 72, 73, 77
van de Hulst, H. C., 20, 24, 51, 76
van den Bergh, S., 286
van Pelt, A., 220, 221, 224, 256
van Woerkom, A. J. J., 157, 169
Vasilevskis, S., 157
Vegard, L., 121, 122, 123, 125
Velghe, A. G., 167, 169
Vestine, E. H., 120, 125
Volders, L., 327
Vorontsov-Velyaminov, B. A., 141, 142
Vyssotsky, A. N., 288

Wahlquist, H. D., 286, 287
Walker, A. G., 312, 313, 318, 323, 328
Walker, M. F., 260, 282, 287, 288
Wallerstein, G., 286
Walsh, D., 317, 327
Waterfield, R. L., 129
Wattenberg, D., 67, 77
Watts, C. B., 154, 170
Weaver, H. F., 162, 170, 198, 202
Weenen, J., 219, 220, 221, 224, 225, 257
Weitbrecht, R. H., 155, 161, 163, 170
Weizsäcker, C. F. von, 309, 310
Wellmann, P., 95, 106, 240, 241, 242, 243, 244, 245, 257
Wenzel, W. A., 325, 327
Whaling, W., 325, 327
Wheelon, A. D., 50, 74, 77
Whipple, F. L., 129, 130, 131, 132, 133, 137, 141, 142, 143, 144
White, H. S., 159, 169, 175, 176, 180, 183, 287
Whitford, A. E., 148, 158, 170, 174, 176, 178, 183, 312, 320, 322, 328
Whitrow, G. J., 311, 317, 326, 328
Wild, P., 324, 327
Willstrop, R. V., 143, 144, 200
Wilson, O. C., 218, 222, 223, 225, 257, 317, 328
Wilson, R. E., 163, 170, 269
Wolf, M., 49, 77
Woolf, N. J., 143

Woolley, R. v.d. R., 3, 15, 17, 18, 22, 23, 24, 50, 65, 75, 175, 183, 208, 257, 272, 288
Wrubel, H., 107, 111
Wurm, K., 216, 239, 240, 242, 256

Yates, G. G., 195, 202
Young, R. K., 58, 65, 75, 77
Yü, C. S., 171, 183

Zanstra, H., 203, 205, 210, 211, 217, 218, 219, 220, 221, 223, 224, 225, 226, 247, 257
Zeiss, C., 67, 94, 96
Zurhellen, W., 52, 77
Zwicky, F., 313, 320, 322, 324, 326, 327, 328

SUBJECT INDEX

Absorption and emission profiles, 217
Absorption components, displacements of, 222, 224
Absorption lines, 203, 206, 214
Absorption of ionizing radiation, 211, 220
Absorption, pure, 206, 208
Abundances, relative, 241, 243
Accelerations, 81, 218
Accretion, 299, 305
Addresses of electronic computers, 109
AG Carinae, 204
AG Pegasi, 227, 228, 229, 230, 233, 234, 239, 243, 253
Airglow, 123
Air-pressure controller, for telescope mirror, 103, 104
All-sky cameras, 117, 118, 119
Angular momentum, 204, 245, 308
Arithmetic unit in computer technique, 108
Artificial satellites, 87, 126, 321
Associations, stellar; accurate motions in, 148, 151
Astrograph, scale-value of a normal, 57
Atmospheric turbulence, 92
Atomic clocks, 85
Atom, schematized, 209
Auroral investigations, techniques, 117
Aurorae, 117–125, 289, 295
Automatic programming of Fortran computer,108
Azimuthal quantum numbers, 234

Babcock's photoelectric template technique, 164
Background density (Schmidt plates), 193
Balancing error (telescope), 188
Balmer decrements, 215, 219, 227, 231, 233, 234, 235, 242, 251, 254
Balmer lines, 209, 210, 214, 215, 219, 230, 233, 245, 247, 252, 254
Beals' hypothesis of an expanding envelope, 217, 218, 221, 222, 223, 224, 225, 254
Becker's iris-photometer, 185, 187
Bell interpretive system, 109, 110
Be-stars, 204, 214, 239, 242, 243, 245, 253, 254
Biermann's theory of cometary tails, 133
Black body, 218, 226
Blanketing effect, 115
Bode's law, 29
Boltzmann factor, 221
Boltzmann-Saha equation, 231
Bondi sphere, 303
Boss General Catalogue, accuracy of positions and proper motions therein, 150, 151, 152, 153
Bowen's fluorescence mechanism, 215, 216, 220
Break-away points, for clusters, 276
Bredichin's theory of cometary tails, 141
Brightness temperature, 223, 228

Cambridge University Observatory; telescope, 92
Carbon-nitrogen cycle, 275
Centring error (telescope), 190
Check-field (Einstein eclipse expeditions), 56
Chemical composition of stars, 266, 285
C/He ratio, 221
Chromosphere, 203, 208
Clock paradox, 78–91
Clusters, ages of, 276
Clusters, development of, 276
Cluster of galaxies, 312
Clusters, galactic, 147, 148
Clusters, globular, 147, 148
Coding (computers), 108
Coherent scattering (in a stationary spherical atmosphere), 212, 245
Colliding galaxies, 315
Collisional de-activation coefficient, 209
Collisional excitation and de-excitation, 206, 208, 209, 210, 211, 235, 237, 238, 239, 240
Colorimetry, 172
Colour, 115, 159, 172, 258, 258–288
Colour: the need for a fourth (infrared) colour, 159
Colour excess, 173
Colour index, 172
Colour-luminosity arrays, 22, 23, 258–288
Colour-luminosity arrays for open clusters, 263
Colour-luminosity arrays for globular clusters, 278
Colour-system transformations, 115
Colour temperature, 180, 204, 208, 218, 223, 236, 251
Conjunction, importance of time; (Uranus), 35
Conjunction, rule for finding, 33
Continuous absorption, 208
Continuous emission, 204, 218
Continuous scattering, 206
Consistency of measurement (Schmidt plates), 190
Consistency (telescopes), 103
Constant-ejection model of a star, 205, 253
Control unit of an electronic computer, 108
Coma Berenices Cluster, 264
Comet Arend-Roland (1956h), 128–144
Comet Arend-Roland, anomalous tail of, 129, 130, 131, 137, 138, 139, 140, 141, 142
Cometary brightness and solar activity, 144
Cometary emissions, 133
Cometary physics, 132
Comet Mrkos (1957d), 128–144
Comets, radio-astronomical work on, 134
Comets, spectra, 133, 134, 135, 144
Computer techniques, 108
Corona, 51, 203
Corrector plate, 99, 185
Cosmical gas-dynamics, 296
Cosmical constant, 318

Cosmological Principle, 312
Cosmological Principle, Perfect-, 312
Cosmology, 311–328
Coudé focus, 99
Crimea work on Be-stars, 245
Curve of growth, 239, 240, 241, 242, 244, 245
61 Cygni group, 270
Cyclic processes, 248

Daytime aurorae, 123
Density of cosmic matter, 318
Density relation (for a uniformly expanding shell), 205, 253
De Sitter Universe, 313
Diffuse nebulae, 211
Diluted black-body radiation, 208, 236
Dilution factor, 204, 210, 216, 217, 221, 229, 235, 242, 243, 247, 250
Diluted stellar radiation, 204, 208, 209, 210, 222, 239, 240, 243
Distance scale, 314
Distribution of galaxies, 323
Dual purpose telescope, 99, 101, 104, 106
Doppler shift, 214, 216, 234, 315
Doppler broadening parameter, 214, 247, 248, 249

Early-type stellar atmospheres, 203–257
Eclipse-plate at Einstein expeditions, 53
Eddington's approximation (source function), 246
Einstein eclipse expedition 1919; 58, 60, 64
— 1922; 58, 61, 65, 68
— 1929; 58, 62, 67, 68
— 1936; 58, 62, 71, 72
— 1947; 58, 63, 72
— 1952; 58, 63, 72
Einstein's field equations, 311
Einstein's light-deflection, 47
Einstein's Principle of Equivalence, 47, 48
Einstein Universe, 313
Electron captures, 209
Electron density, 204, 209, 228, 229, 230, 235, 245, 252
Electron pressure, 207
Electron scattering, 209, 223, 224, 244, 252
Electron shell, 222
Electron temperature, 210, 211, 220, 235, 242, 248
Electronic computers (EDSAC, etc.), 107, 110, 185, 187
Electronic computers: for setting up observing programmes, for data-reduction and catalogue compilation, 154, 160
Electronic computers in stellar photometry, 189
Element formation (α-β-γ theory), 324
Ellipsoidal distribution of stellar velocities, 5
Elliptic velocity distribution, tilt of the axis of the, 18
Emission lines, 203, 204, 205, 206, 207, 208, 210, 217, 219, 222, 224, 229, 240
Emission measure of intergalactic hydrogen, 321
Emission to absorption intensity-ratio, 218
Energy balance, 211
Energy, gravitational, 306
Energy, magnetic, 306
Envelopes of stars, 204, 205, 208, 212, 214, 217, 218, 222, 225, 230, 241, 243, 244, 245, 248, 254, 275
Envelopes, sizes and densities of transparent, 225
Equal galactic period, condition of, 21
Equi-energy source, 182
Equipartition of energy, 23
Equivalent widths, 210, 212, 220, 225, 226, 241, 251, 253
Evolutionary history of the stars, 258–288
Excitation temperature, 245
Expansion factor of the Universe, 312
Expansion, from the stellar surface, 204, 214, 227, 229
Explosive models of the Universe, 319
Extinction, 113, 115

Fabry-Perot interferometer, in auroral research, 121, 122
Fehrenbach prism, 94
Ferraro-Alfvén effect, 305
Field errors, (Schmidt plates), 195
Filter-photography of comets, 143, 144
Filters, 172, 204
Flexure, telescope, 184
Flow diagram, 110
Fluorescence, effects and mechanism, 204, 206, 208, 210, 214, 215, 216, 217, 218, 219, 246, 252, 254, 255
Forbidden lines, 203, 211, 220, 236
Force-fit method of determining cluster distances, 271
Formation of galaxies, 323
Formation of stars, 296–310
Formation of the elements, 324
Frames of reference, inertial and non-inertial, 80, 88
Fraunhofer lines, central intensities of, 208

Galactic magnetic field, 296
Galactic orbits and colour-luminosity arrays, 22
Galactic orbits of stars, 8, 22, 272
Galaxies, age distribution, 321
Galaxies, clusters of, 312
Galaxies, distribution of, 323
Galaxies, formation of, 323
Galaxies, relaxation time, 23
Galaxy counts, 313
Gamma Cassiopeiae, 242, 243
Gas dynamics, 3
Gaseous nebulae, 203, 210, 215, 217, 235, 246, 247
Gaunt factor, 231, 236
Globular clusters, colour-luminosity arrays, 278
Gold point (high-temperature standard), 182
Granularity (Schmidt plates), 191
Gravitational contraction, 285
Gravitational field, 47, 84, 244, 247
Gravitational potential, 81
Gravitational red-shift, 84
Gregorian focus, 99

H I regions, 205, 229, 298
H II regions, 205, 211, 229, 236, 304

Hamburg Observatory; telescope, 94
Harvard Nebular Series, 205
Haute Provence, Observatoire; telescope, 92, 98
He/H ratio, 221
ζ Herculis group, 270
Hertzsprung-Russell diagram, tracks of evolution in the, 115
Hiltner's seeing-compensated polarization technique, 166
Homogeneous stars, convective core, 275
Hubble law, 313
Hyades, 268
Hydrogen atom, schematized, 208
Hydrogen atoms, distribution among sub-levels, 234
Hydrogen content of stars, 266
Hydrogen nebulae, 211
Hydromagnetic waves, 305
Hydrostatic equilibrium, 203, 204

Idealized atom, 240, 248
Image structure (Schmidt plates), 194
Induced emissions, 231, 253
Inertial frame, 80
Infrared spectra, 219
Inglis-Teller formula, 288, 229
Input (computer), 108
Instrumental colour system, transformation of the, 113
Intercombinations, 216
Interference filters, 144, 160, 161
Intergalactic matter, 318, 320
Interlocking, 208
Interpretive systems in computer programming, 109
Interstellar polarization, 166
Interstellar reddening, 204, 218, 228, 251
Intrinsic reddening in P Cygni, 204
Ionization, 208, 209, 210, 211, 217, 218, 219, 221, 226, 231, 232, 235, 239, 248, 321
Ionization equation, 211
Ionization from excited states, 210
Ionization from ground state, 208
Ionization potential, 217, 218, 221
Ionization temperature, 221
Ionized region, extent of, 210, 211
Ionizing continuum, 208, 248
Iris-photometers, 185, 188
Isaac Newton Telescope, 99, 100

Keplerian rotation field, 305
Kepler's Third Law, 204
Kelvin-Helmholtz contraction, 310
Kelvin-Helmholtz time-scale, 299
Kirchhoff's law, 206
K-term, 314

Late-type stars with emission lines, 229
Leakage effect, 214, 221, 234, 248, 253
Lick Observatory (Einstein expedition), 48
Light receivers, 172
Limb-brightening, 246
Limb-darkening law, 254
Limb, of stars, 204, 205
Linearity (telescopes), 103
Linear-flow approximation, 206
Line profiles, 115, 203, 213, 218, 221, 230, 241, 250, 253, 254, 255
Line-scattering in an extended moving shell, 212
Liouville's theorem, 4
Lorentz transformation, 78
Luminosity, 258–288
Luminosity class, photoelectric determination of, 160
Luminosity function, 285
Lyman continuum, 205, 210, 218, 228, 244, 247, 248
Lyman limit, 205, 210, 211, 228, 246, 252
Lyman lines, 205, 210, 211, 219, 226, 242

Machine language (computer), 108
Magellanic Clouds, 289, 290, 291, 294
Magnetic clouds, 296
Magnitude scales by photoelectric photometry, 201, 202
Magnitudes and multicolours; proposal for observation of the 70,000 brightest stars, 158
Magnitude transfer in stellar photometry, 198, 200
Main mirror, 99
Main sequences of stars; slope and zero point, 261, 275
Masers (microwave amplifiers), 317
Mass–hydrogen-content diagram, 266
Maxwellian velocity distribution, 211
Mechanical excitation, 217
Memory (computer), 108
Meson beam, 83
Metastable states, 204, 215, 229, 239, 240, 242, 243, 245
Milky Way, visual representation, 289–295
Milne-Eddington model, 115, 206
Mirror cell, 95
Model atmospheres, 207, 208, 218
Moonwatch-type telescopes, 126
Mount Palomar Observatory; telescope, 99
Moving envelope, 248
Multiple-colour photometry, 171, 174, 176

Nebulae, 203, 204, 210, 211, 215, 217, 235, 246, 247
Nebular decrement, 226, 234, 251
Nebulium, 211
Neptune, rediscovery of, 25
Ne-stars, 51
Neutral gas, 300
Neutral hydrogen, 21 cm-line, 317
Night-plate at Einstein eclipse expeditions, 53
Non-coherent scattering, 213, 247, 249
Novae, shells and spectra, 203, 204, 217, 252, 253, 282
Nucleogenesis, 325

Occultation effect, 214, 222, 224, 226, 233, 241, 242
Of-stars, 208, 215, 216, 217, 224

Omega Centauri Cluster, 289, 294
Oort constants of galactic rotation, 5
Opaque region, extent of, 254
Optical alignment of telescopes, 93
Optical depth, horizontal, 205
Optical depth, radial, 245
Optical depth, tangential, 227
Optical observations of the Russian satellite, 126
Optimum density (photometric measurements), 198
Output (computer), 108

P Cygni stars, 203, 204, 214, 215, 216, 217, 218, 219, 226, 227, 228, 229, 230, 233, 234, 239, 240, 243, 245, 250, 251, 253, 254, 255
Pannekoek's ionization equation, 228, 237
Parallaxes of stars, 162
Paschen/Balmer ratio, 226
Paschen decrement, 233
Paschen lines, 215, 219, 229, 251, 252
Perfect cosmological principle, 312
Permanence of very open clusters, 20
Petzval surface, 184
Plasma, 298, 301
Pic du Midi Observatory (telescope), 92
Pickering series, 251
Pleiades, 262
Phase-lag in velocity, of expanding envelopes, 222
Photocells, 173
Photoelectric observations, reduction of, 112
Photoelectric observatories, the need for, 159
Photoelectric temperatures, 219, 221
Photo-excitation, 226, 231
Photographic auroral observations, 117
Photographic emulsion and processing, 51
Photo-ionization, 209, 226, 231, 232, 235, 239
Photometric data, intrinsic scatter in, 180
Photometric systems, 172
Photomultipliers, applied to auroral research, 123
Photomultiplier tubes, the need for more efficient infrared-sensitive, 159
Planck function, 206, 207
Planetary nebulae, 204, 205, 208, 211, 212, 216, 217, 218, 219, 221, 230, 233
Polarization, 223
Population by cascade, 235
Population of atomic levels, 208
Positions, accuracy of meridian circle, 154
Positions, accuracy of the GC, 151
Positions, accuracy of wide-angle plates, 156, 157
Potsdam Observatory (telescope), 96
Praesepe, 260
Prime focus, 99
Primeval gas cloud, 285
Principle of Equivalence, 47, 89
Profiles of lines, 111
Programme library, 108
Programming of electronic computers, 108
Programming system SOAP, 110
Prominences, 204
Proper motions, accuracy from wide-angle plates, 156, 157
Proper motions, accuracy of the GC, 150, 151, 152, 153
Proper time, 88
Proton-proton chain, 275
Proto-star, 296, 305
Pulsation, 254
Purkinje effect, 172

RR Lyrae stars, 282
Radial velocities, photoelectric determination of 164
Radial-velocity photometer, 164
Radial velocity, surfaces of equal, 254
Radiation pressure, 205, 244, 247
Radiation through a foggy atmosphere, 206
Radiation transfer, 209, 245, 246, 247, 248
Radiative equilibrium, 203, 206, 246
Radiative transfer in a moving shell, 205, 247
Radio astronomy, 317
Radio-echo, auroral work, 124
Radio observations of comets, 134
Radio sources, 315
Ratio of Hα to Hβ, 252
Rayleigh's criterion, 310
Recombination, 208, 209, 210, 215, 216, 219, 220, 222, 226, 231, 234, 235, 236, 246
Recombination to excited states, 208, 210
Recombination to the ground state, 210
Red-shift, gravitational, 84
Red-shift, of galaxies, 313
Residual intensity in a spectral line, 208, 209
Ring nebula in Lyra, 204
Robertson-Walker metric of the Universe, 312
Rosseland cycle, 208, 210, 216, 221, 242, 247
Rotating envelope, 204, 214, 217, 243, 244, 254
Rotational broadening, 204, 245
Rotational velocity, 285
Rottenberg's line scattering mechanism, 217
Royal Greenwich Observatory (telescope), 94

Sartorius irisphotometer, 185, 186
Satellite clock, 85
Satellites, artificial-, 87, 126, 321
Scale-value, Einstein eclipse observations, 53, 55
Scattering, 206, 208, 210, 212, 213, 214, 218, 219, 224, 245, 247, 248, 249
Scale height of stellar atmospheres, 204
Schilt photometer, 194
Schmidt telescope, 99, 184–202
Schuster-Schwarzschild approximation, 246
Schuster-Underhill mechanism, 206, 210
Seeing, 200
Seidel aberrations, 93
Selection effects (in velocities of late type stars), 18
Selectivities, in emission-line stars, 206, 215, 216, 217, 224
Self-absorption, 217, 218, 219, 224, 226, 227, 234, 240, 241, 243, 250, 251, 253, 254
Serrurier telescope-tube construction, 99
Se-stars, 251
Shells (and rings, surrounding emission-line stars), 203, 204, 208, 210, 211, 212, 213, 219, 221, 234, 239, 242, 244, 245, 248, 253
Six-colour photometry, 174, 176
Sirius group, 270

Solar corona, 51, 203
Solar disc, 208
Solar spectrum, 208
Source function, 209, 213, 246, 249
Southern-hemisphere observatory, 156
Space, 82
Space motions of stars, 165
Space-time, 82
Space travel in a gravitational field, 84
Space travellers, 85, 86
Spatial curvature, 312
Spectral gradient, 208
Spectral types, photoelectric determination of, 160
Spectroheliograms, 208
Spectrophotometry, 172, 219
Spectroscopy, of aurorae, 122
Speed, of response (telescopes), 103
Spherical shell, stationary, 245
Spontaneous transition rate, 209
Stability (telescopes), 102
Standard Praesepe-Pleiades relationship, 283
Star-fields at Einstein-eclipse observations, 56
Star formation, 296 310
Stark broadening of hydrogen lines, 115
Stars, the 70,000 brightest, 147, 149
Statistical equilibrium, 205, 209, 224, 226, 230, 231, 235, 236, 239, 240, 243, 248, 249, 250, 255
Steady-state theory of the Universe, 3, 313
Stebbins-Whitford effect, 322
Stellar atmospheres, 203, 204, 206, 208, 218
Stellar dynamics, 3
Stellar photometry, use of punched tapes, 189
Stellar rotation, 239, 305
Stellar temperatures, 205, 206, 210, 218, 220
Stellar velocities, 15
Stiction in telescope bearings, 97
Stimulated captures, 236
Stimulated emissions, 209, 250, 251
Stratification, 212, 214, 215, 217, 218, 223, 224, 245
Strömgren's narrow-band filter technique, 160
Sub-levels, 204, 234, 235
Subroutine of electronic computing work, 113
Subsonic flow, 303
Sun, 208
Supergiants, 203, 205, 206, 207, 252
Supersonic flow, 303
Surface gravity, 221
Surface temperature of stars, 277
Synchronization of clocks, 79
T Tauri variables, 284
Telescopes, design of large, 92
Temperature compensation (Schmidt telescope), 186
Tidal effects (Wolf-Rayet stars), 222
Time, 82
Time of relaxation, of galaxies, 23
Theory of Relativity, 47, 78
Thermal distortion, 184
Thermodynamic equilibrium, 203, 208, 209, 220, 221
Thomson scattering, 205, 206, 223, 224
Three-body recombinations, 236
Three-colour (U, B, V) photometric system, 174
Turbulence, 203, 229, 244, 309

Ultra-violet continuum, 218
Uniform-expansion stellar model, 229, 230
Universe, structure of, 311
Uranus, 25
Uranus, equations of condition (Neptune discovery), 42
Uranus, heliocentric conjunction, 30

V 444 Cygni, 215, 221, 222, 223, 224, 246
Veiling of stellar absorption lines, 244
Velocity dispersion of stars, 278
Velocity gradients, 218, 248, 249, 250, 253
Visual auroral observations, 117
Visual Milky Way, 289–295

WC sequence, 221
Whipple's comet theory, 131, 133, 142
White dwarfs, 282
Wilson's 'transit-time' hypothesis, 222
WN sequence, 221
Wolf-Rayet stars, 203, 204, 212, 215, 217, 218, 219, 220, 221, 223, 250, 251, 253
Wolf-Rayet stars, densities of, 221
Wolf-Rayet stars, sizes of, 221
'World-map' of the Universe, 323
'World-picture' of the Universe, 323
W Ursae Majoris eclipsing binaries, 285

Zanstra's fluorescence hypothesis, 203
Zeeman splitting, 227
Zeta Tauri, 242
'Zero-age' sequence, 284

PART I

DYNAMICS
INSTRUMENTS
GEOPHYSICS
SOLAR SYSTEM

The Dynamics of Stars in the Neighbourhood of the Sun

R. v. d. R. Woolley, F.R.S., Astronomer Royal
Royal Greenwich Observatory, Herstmonceux Castle, Sussex

1. Introduction

A certain amount is known about the positions and motions of nearby stars: the knowledge becomes more and more fragmentary as the distance of the stars from the Sun increases. Nevertheless an enormous body of information of a partial and inaccurate character exists, from which statistical inferences may be drawn. It is the business of Stellar Dynamics to attempt to reconcile this body of observed material with some orderly view of the forces which influence stellar motions, in the hope of establishing data about the dimensions and age of the Galaxy.

Besides the stars themselves, observation reveals the presence of dust and gas in such large quantities as to account for an appreciable fraction of the mass of the Galaxy: and theory gives a time of relaxation of the stellar motions, through stellar encounters, of order 10^{12} years, and so much greater than the supposed age of the Galaxy of 10^{10} years. It is for this reason that the present motions of the stars are by no means random motions, but are no doubt still very markedly influenced by initial conditions. If the stars were formed out of clouds of gas, these initial conditions, the initial vectors of the motion of each star when first formed, are doubtless to be looked for in a theory of motion of gaseous clouds rather than in a "Newtonian" theory of point masses moving in a gravitational field. For all this, the stars once formed move as "particles", and the only role of the dust and gas is to modify the gravitational field. There is therefore some sense in studying the present movements of stars as particle-movements in a gravitational field, and this is the traditional field of Stellar Dynamics. We merely recognize that Gas Dynamics may have some connection with our initial conditions: we remember that our attraction is not solely due to the stars themselves: and apart from that, we dismiss Gas Dynamics as a separate subject.

2. Steady-state Theory

2.1. There is no reason to regard the motions of stars in the solar neighbourhood as a steady state—quite the reverse. The distribution in space is irregular and the supposed time of relaxation much greater than the life of the system since its inception. Nevertheless, a mathematical investigation of a steady state or quasi steady state is instructive.

The standard procedure is as follows, (see, e.g. Smart, 1938; p. 321, etc.). Let

$$f(u, v, w, x, y, z, t)\, \mathrm{d}u\, \mathrm{d}v\, \mathrm{d}w\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$$

be the number of stars in unit volume with velocities between u and $u+\mathrm{d}u$ etc., and positions between x and $x+\mathrm{d}x$ etc. Then if the motions are entirely governed by a

gravitational potential V which is a function of x, y and z (i.e. is independent of t which would not be true if collisions were considered) we have LIOUVILLE's theorem

$$\frac{Df}{Dt} = 0$$

The number of stars per unit volume is

$$\nu = \iiint f(u, v, w, x, y, z, t)\, du\, dv\, dw,$$

and the system is said to be in a *steady state* if

$$\frac{\partial \nu}{\partial t} = 0$$

so that the function f satisfies

$$\frac{\partial f}{\partial t} = 0 \tag{2.11}$$

This last relation can be developed in various ways, but we proceed immediately to a form suitable for discussing stars moving in an outer portion of a galaxy, that is to say a body possessing symmetry about an axis of rotation.

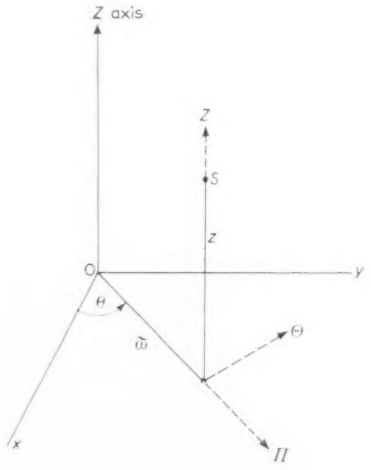


FIG. 1.

Let (ϖ, θ, z) be the cylindrical co-ordinates of a star S as shown in Fig. 1, and let the components of stellar velocity be Π outwards from the origin, Θ perpendicular to the radius vector and in the plane of rotation passing through the origin, and Z perpendicular to the plane of rotation. If the function f and the potential V are symmetrical about the axis of rotation, (2.11) becomes (cf. SMART, 1938, p. 338):

$$\Pi\frac{\partial f}{\partial \varpi} + Z\frac{\partial f}{\partial z} + \left(\frac{\partial V}{\partial \varpi} + \frac{\Theta^2}{\varpi}\right)\frac{\partial f}{\partial \Pi} - \frac{\Pi\Theta}{\varpi}\frac{\partial f}{\partial \Theta} + \frac{\partial V}{\partial z}\frac{\partial f}{\partial Z} = 0 \tag{2.12}$$

This is satisfied by

$$f = F(\Pi^2+\Theta^2+Z^2-2V, \varpi\Theta) \qquad (2.13)$$

2.2. We now consider a particular form of (2.13) suggested by the statistics of stellar velocities, the *ellipsoidal distribution*, or

$$F = \exp -j^2\{\Pi^2+\lambda^2(\Theta-\Theta_0)^2+Z^2-2V_1\} \qquad (2.21)$$

which satisfies (2.12) if

$$\begin{aligned} \lambda^2 &= 1+k_2\varpi^2 \\ \lambda^2\Theta_0 &= -k_1\varpi \\ V_1-V &= \tfrac{1}{2}\lambda^2\Theta_0^2 \end{aligned} \qquad \ldots\ldots (2.22)$$

where j, k_1 and k_2 are constant.

Since the star density is given by

$$\nu = \iiint f \,\mathrm{d}\Pi\,\mathrm{d}\Theta\,\mathrm{d}Z,$$

we have from (2.21)

$$\nu = \text{const} \times \lambda^{-1}\exp(2j^2V_1). \qquad \ldots\ldots (2.23)$$

Equation (2.21) also gives the mean values of Π, $\Theta-\Theta_0$ and Z, taken without regard to sign, as

$$\bar{\Pi} = \int_0^\infty \Pi e^{-j^2\Pi^2}\,\mathrm{d}\Pi \Big/ \int_0^\infty e^{-j^2\Pi^2}\,\mathrm{d}\Pi$$

and we have in fact

$$j\bar{\Pi} = \frac{2}{\sqrt{\pi}}, \quad j(\overline{\Theta-\Theta_0}) = \frac{2}{\sqrt{\pi}}\frac{1}{\lambda}, \quad j\bar{Z} = \frac{2}{\sqrt{\pi}}$$

and the ratios of the mean velocities are

$$\bar{\Pi} : \overline{\Theta-\Theta_0} : Z :: \lambda : 1 : \lambda \qquad \ldots\ldots (2.24)$$

those in the radial co-ordinate $\bar{\Pi}$ and co-ordinate perpendicular to the galactic plane $\bar{Z}$ being equal, and greater than that in the tangential co-ordinate $\overline{\Theta-\Theta_0}$ if k_2 is positive and λ greater than unity, which we will show should be the case in the solar neighbourhood.

The constants k_1 and k_2 can be connected with the Oort constants A and B which are accessible to observation. We have in fact*

$$\begin{aligned} A &= \frac{1}{2}\frac{\mathrm{d}\Theta_0}{\mathrm{d}\varpi}-\frac{1}{2}\frac{\Theta_0}{\varpi} \\ B &= \frac{1}{2}\frac{\mathrm{d}\Theta_0}{\mathrm{d}\varpi}+\frac{1}{2}\frac{\Theta_0}{\varpi} \end{aligned} \qquad \ldots\ldots (2.25)$$

* Θ is measured in the sense in which longitude increases: this happens to be opposite to the sense in which the Galaxy is rotating. If V is the velocity, considered positive, and R the distance from the galactic centre, $A = \frac{1}{2}V/R-\frac{1}{2}\mathrm{d}V/\mathrm{d}R$.

so that

$$A = k_1 k_2 \varpi^2 \lambda^{-4}, \; B = -k_1 \lambda^4$$

and

$$k_1 = -\lambda^4 B, \; k_2 = -A/(B\varpi^2).$$

A is observed to be positive and B to be negative: hence both k_1 and k_2 are positive and λ is greater than unity.

Notice that $\lambda^2 = 1 + k_2 \varpi^2 = (B - A)/B$.

The sign of k_2 is also indicated by theory; if Θ_0 is identified with, or is very nearly equal to, the circular velocity, which is Θ_c, we have

$$\Theta_c^2 = \mu \varpi^{-1}$$

where μ is the effective attracting mass of the Galaxy multiplied by the gravitational constant, or

$$\mu = \frac{4}{3} \pi \Gamma \bar{\rho} \varpi^3$$

in which $\bar{\rho}$ is the effective mean density.

Similarly $\mathrm{d}\mu/\mathrm{d}\varpi = 4\pi\Gamma\rho\varpi^2$ where ρ is the effective local density (and exactly equal to the local density in the case of spherical symmetry). Then

$$k_2 = -\frac{A}{B\varpi^2} = \frac{1}{\varpi^2} \cdot \frac{3(1 - \rho/\bar{\rho})}{1 + 3\rho/\bar{\rho}}$$

which is positive unless $\rho/\bar{\rho} > 1$, which is not likely to be the case in the outer parts of a galaxy.

Steady-state theory, then, supposes $\lambda > 1$ and equality between the mean speeds in the radial and z directions. This is not indeed found in the observations. The observed velocity distribution for any particular class of star can be represented roughly by

$$n(u, v, w) = \text{const} \times \exp -\{j_1^2 u^2 + j_2^2 v^2 + j_3^2 w^2\}$$

where the u, v, w axes are close to (but not identical with) the Π, Θ, Z axes, but the quantity j_1 is not found to be equal to j_3. We do find $j_1 < j_2$ as steady-state theory suggests, but j_3 is nearer to j_2 than to j_1.

This is probably not a major difficulty and, no doubt, arises out of the fact that relaxation is as yet nowhere near complete and no connexion has been established by collisions between the energy of the z motion and energy in the plane. If the motions in the plane and perpendicular to the plane were quite independent, we should have two separate energy equations

$$I_1 = \Pi^2 + \Theta^2 - 2V_1, \; I_2 = Z^2 - 2V_2$$

leading to

$$F_1 = \exp -j_1^2\{\Pi^2 + \lambda^2(\Theta - \Theta_0)^2 - 2V_1\}$$
$$F_2 = \exp -j_2^2\{Z^2 - 2V_2\}$$

the second of these equations being usually assumed to be correct when motion perpendicular to the galactic plane is discussed. The separation of the energy equations breaks the relation of equality between $\overline{\Pi}$ and $\overline{Z}$ given by equation (2.24).

Steady-state theory gives a relation between Θ_0 and Θ_c, the circular velocity. Θ_c is given by

$$\Theta_c^2 = -\varpi \frac{\mathrm{d}V}{\mathrm{d}\varpi}$$

and from equations (2.22) we can derive

$$\Theta_0^2 = \frac{1}{2}\varpi \frac{\mathrm{d}}{\mathrm{d}\varpi}(\lambda^2\Theta_0^2).$$

From equation (2.23) we have

$$\ln \nu = \text{const} - \ln \lambda + 2\gamma^2 V_1$$

whence

$$\frac{1}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\varpi} = -\frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}\varpi} + 2j^2\frac{\mathrm{d}V}{\mathrm{d}\varpi} + j^2\frac{\mathrm{d}}{\mathrm{d}\varpi}(\lambda^2\Theta_0^2).$$

Lastly

$$\frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}\varpi} = \frac{1}{\varpi}\frac{\lambda^2-1}{\lambda^2}.$$

Hence Θ_0 is given by

$$2j^2(\Theta_0^2 - \Theta_c^2) = \frac{\varpi}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\varpi} + \frac{\lambda^2-1}{\lambda^2}$$

the quantities on the right-hand side being, in principle, accessible to observation (through the radial stellar density gradient and ratio of the axes of the velocity ellipse).

If we write $\Theta_0 = \Theta_c - S$ and neglect S^2, and use $\Theta_0 = \varpi(B-A)$

$$S = \frac{1}{4j^2(A-B)}\left(\frac{1}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\varpi} + \frac{\lambda^2-1}{\varpi\lambda^2}\right). \qquad \ldots\ldots (2.26)$$

This means that Θ_0, the mean velocity for a group of stars characterized by a particular value of j^2, that is a particular dispersion in velocity, differs from the circular velocity by an amount inversely proportional to j^2, or directly proportional to the mean square (dispersion) velocity. There is some justification for this in the available data.

The successes of steady-state theory are merely that if an ellipsoidal velocity distribution (2.21) is *assumed*, the theory asserts a ratio of the axes (2.24), which can be reconciled with observation as far as the Π and Θ axes are concerned, and explained away for the Z axis: and that it asserts through (2.26) a dependence of solar motion on velocity dispersion which is useful. Limited as these successes are, it is perhaps rather surprising that the theory is of anything but academic interest. The

spatial distribution of the stars is anything but regular, except that *on the whole* there is a decrease of density outwards from the centre: and we may ask whether there is any reason to expect greater regularity in the velocity distribution than we find in the space distribution.

3. Galactic Orbits of Stars

3.1. We now turn to another approach to the analysis of stellar velocities, due in principle to the researches of Lindblad, and consider the orbits of individual stars moving in the gravitational field of the galaxy as a whole. The treatment in the preliminary section (3.1) is equivalent to that given, for example, in Smart (1938, Chapter 12), except that in the development given here the co-ordinate axes are non-rotating.

We leave motion perpendicular to the galactic plane out of account and consider only motion in the galactic plane. If ϖ is the distance to the centre of the galaxy and $\boldsymbol{u} = \varpi^{-1}$ the (component in the galactic plane of the) attracting force may be written $\boldsymbol{u}^2\mu(\boldsymbol{u})$. The equation of motion of the star is then

$$\frac{\mathrm{d}^2\boldsymbol{u}}{\mathrm{d}\theta^2} + \boldsymbol{u} = \frac{\mu(\boldsymbol{u})}{h^2}$$

where h is a constant (since the force may be assumed central). If we may write, with sufficient accuracy,

$$\mu(\boldsymbol{u}) = \mu_0 + (\boldsymbol{u} - \boldsymbol{u}_0)\mu' \qquad \ldots . (3.11)$$

the equation for the orbit is

$$\boldsymbol{u} = \varpi_1^{-1}(1 + e\cos n\theta)$$

where

$$n^2 = 1 - \mu'/h^2 \quad \text{and} \quad \varpi_1 = (h^2 - \mu')/(\mu_0 - \boldsymbol{u}_0\mu')$$

Let μ_1 be the value of $\mu(\boldsymbol{u})$ when $\varpi = \varpi_1$, that is to say let

$$\mu_1 + (\boldsymbol{u} - \varpi_1^{-1})\mu' = \mu_0 + (\boldsymbol{u} - \boldsymbol{u}_0)\mu'$$

Then

$$\mu_1 = h^2/\varpi_1$$

and the equation of the orbit is

$$\boldsymbol{u} = \frac{\mu_1}{h^2}(1 + e\cos n\theta)$$

If Θ_c is the circular velocity at ϖ, we have

$$\mu_1 = \varpi_1\Theta_c^2$$

Accordingly

$$h = \varpi_1\Theta_c$$

whatever the eccentricity of the orbit.

If the eccentricity is so small that its square may be neglected (a restriction that is in any case suggested by the linear form adopted for $\mu(\boldsymbol{u})$), the motion (3.11) may be regarded as an elliptic epicycle centred on a point moving with the circular velocity.

In Fig. 2, O is the centre of attraction and C a point moving in a circle centred on O with radius ϖ_1, the velocity of C being Θ_c. The angle made by OC with the fixed direction OX is θ. Let P be a point whose co-ordinates $(\xi_1\eta)$ in rectangular axes whose origin is the instantaneous position of C and such that the axis of ξ is OC produced, and let

$$\left.\begin{aligned} \xi &= -e\varpi_1 \cos n\theta \\ \eta &= \frac{2e}{n}\varpi_1 \sin n\theta \end{aligned}\right\} \qquad \ldots (3.12)$$

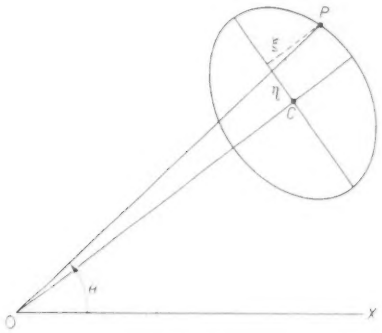


Fig. 2.

Then the angle POC is $\sin^{-1}\eta/(\varpi_1+\xi)$, or, neglecting e^2,

$$\text{angle } POC = \frac{2e}{n}\sin n\theta.$$

The length OP is $(\varpi_1+\xi)\sec POC = \varpi_1+\xi$, again neglecting e^2, so that

$$(OP)^{-1} = \varpi_1{}^{-1}(1+\xi/\varpi_1) = \varpi_1{}^{-1}(1+e\cos n\theta) = \boldsymbol{u}$$

Lastly

$$\dot\theta = \Theta_c/\varpi_1 + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{2e}{n}\sin n\theta\right),$$

giving

$$\dot\theta\,(1-2e\cos n\theta) = \Theta_c/\varpi_1$$

or, neglecting e^2,

$$\dot\theta = h\boldsymbol{u}^2.$$

Accordingly the point P satisfies equation (3.11). The use of an epicyclic description of the motion of the star offers advantages over the use of the galactic orbit (3.11),

because ϖ_1 and Θ_c are not known accurately and because it is our object to analyse the statistics of *differences* between individual stellar velocities and the circular velocity, and these differences are closely related to the epicycles.

The velocity (Π, Θ) of P in *fixed* axes whose origin is the instantaneous position of C is given, neglecting e^2, by

$$\left.\begin{aligned} \Pi &= \dot{\varpi} = \Theta_c e n \sin n\theta \\ \Theta &= \varpi\dot{\theta} = h\boldsymbol{u} = \Theta_c(1 + e\cos n\theta) \end{aligned}\right\}. \qquad \ldots\ldots (3.13)$$

Both θ and Θ_c are measured in the direction in which longitude increases.

In discussing the motions of a number of stars it is necessary to use axes which are common to the stars examined. The system which we will use is a system of rectangular axes whose directions are fixed, the x axis pointing away from the galactic centre, the y axis in the direction in which rotation takes place, and the z axis perpendicular to the galactic plane. The origin moves with constant velocity in the direction of the y axis, and does so with the circular velocity. If the velocities in this system are (u, v, w) then the velocity $(0, 0, 0)$ is the circular velocity at the origin of co-ordinates.

If we write V for the circular velocity in the direction in which galactic rotation takes place, and R for the distance to the centre, then the Oort constants A and B are defined by (2.25), or, with $V = -\Theta_c$,

$$2A = -\frac{\mathrm{d}V}{\mathrm{d}R} + \frac{V}{R}$$

$$2B = -\frac{\mathrm{d}V}{\mathrm{d}R} - \frac{V}{R}.$$

To the first order in e, $u = \Pi$, but v is measured relative to the circular velocity at P (for a star near the origin of the (u, v, w) system), while $\Theta - \Theta_c$ is measured relative to the circular velocity at C. Also the sign is changed. Hence

$$-v = (\Theta - \Theta_c) - \xi\frac{\mathrm{d}\Theta_c}{\mathrm{d}\varpi}.$$

Accordingly,

$$\left.\begin{aligned} u &= -neR(A-B)\sin n\theta = -Ven\sin n\theta \\ v &= -2BeR\cos n\theta = \tfrac{1}{2}Ven^2\cos n\theta \end{aligned}\right\}. \qquad \ldots\ldots (3.14)$$

The co-ordinates of the epicentre C, relative to P as origin, may be defined as x (radial, positive outwards) and y (tangential, positive in the direction in which rotation takes place). Then

$$\left.\begin{aligned} x &= eR\cos n\theta \\ y &= \frac{2eR}{n}\sin n\theta \end{aligned}\right\}. \qquad \ldots\ldots (3.15)$$

Neglecting e^2, the quantity n^2 is given by $n^2 = 4B/(B-A) = 4/\lambda^2$, the quantity λ

having been defined in section 2. To show this we note that $\mu_1 = \varpi_1\Theta_c^2 = RV^2$, so that

$$\mu'\left(\equiv \frac{\mathrm{d}\mu}{\mathrm{d}\boldsymbol{u}} \right) = -R^2\frac{\mathrm{d}\mu}{\mathrm{d}R}$$

or

$$\mu' = -R^2V^2\left(1+\frac{R}{V}\frac{\mathrm{d}V}{\mathrm{d}R}\right).$$

But

$$h^2 = \varpi_1^2\Theta_c^2$$

Hence, neglecting e^2,

$$n^2 = 1-\mu'/h^2 = 1-\mu'/(R^2V^2) = 2\left(1+\frac{R}{V}\frac{\mathrm{d}V}{\mathrm{d}R}\right) = \frac{4B}{B-A}. \quad \ldots . (3.16)$$

With this value of n, we have, from (3.15),

$$\frac{x}{v} = -\frac{y}{u} = -\frac{1}{2B}. \quad \ldots . (3.17)$$

By (3.14),

$$e^2R^2 = \frac{u^2}{n^2(A-B)^2}+\frac{v^2}{4B^2}$$

or

$$4B^2e^2R^2 = \frac{u^2}{\lambda^2}+v^2. \quad \ldots . (3.181)$$

This is equivalent to

$$n^2V^2e^2 = u^2+\lambda^2v^2 \quad \ldots . (3.182)$$

using (3.6) and $V = R(A-B)$.

The epicentre C is inside the present position P of the star if x is negative, and, since B is negative, this occurs, by (3.17) when v is negative and the star lags behind the circular velocity. If v is positive the epicentre is outside the present position of the star. In either case the closest approach to the centre of the Galaxy is the distance eR inside the epicentre and $eR-x$ inside the present position. The perigalactic distance is therefore

$$R+\frac{1}{2B}\left[\sqrt{\left(\frac{u^2}{\lambda^2}+v^2\right)}-v\right]. \quad \ldots . (3.19)$$

All the formulae in this section are only valid if e^2 can be neglected, that is to say when u/V and v/V are small.

3.2. Let $\nu(u, v)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ be the number of stars in an elementary volume

$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ at the origin, which velocities between u and $u+du$ and between v and $v+\mathrm{d}v$. The epicentres of their orbits lie in a volume $\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z$ at (x, y) where

$$R^2\mathrm{d}x'\mathrm{d}y' = (R+x)^2\mathrm{d}x\mathrm{d}y$$

and the eccentricities and phases lie between e and $e+\mathrm{d}e$ and between ψ and $\psi+\mathrm{d}\psi$ where

$$u^2+\lambda^2v^2 = n^2V^2e^2$$
$$\lambda\mathrm{d}u\,.\,\mathrm{d}v = e\mathrm{d}e\,.\,\mathrm{d}\psi\,.\,n^2V^2.$$

Then if $\rho\,(x, y)$ be the number of epicentres per unit volume and if $p(e^2)\mathrm{d}(e^2)$, $p(\psi)\mathrm{d}\psi$ are the fractions of the squared eccentricities and phases which lie between e^2 and $e^2+\mathrm{d}(e^2)$ and between ψ and $\psi+\mathrm{d}\psi$ respectively,

$$\nu(u, v)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}y = \rho(x, y)\,p(e^2)p(\psi)2e\,\mathrm{d}e\,\mathrm{d}\psi\,\mathrm{d}x'\,\mathrm{d}y'$$

and

$$\tfrac{1}{2}n^3V^2R^2\nu(u, v) = 2(R+x)^2\rho(x, y)p(e^2)p(\psi). \qquad \ldots\ (3.21)$$

The quantity $p(\psi)$ may be calculated by supposing that the distribution of phases at a given *place* is the same as the distribution of such phases in *time* as the stars describe their orbits. For each star

$$\dot{\theta} = hu^2 = \mu_1{}^2/h^3\,.\,(1+e\cos n\theta)^2$$

or, since $\psi = n\theta$,

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{n\mu_1{}^2}{h^3}(1e+\cos\psi)^2.$$

Accordingly $\mathrm{d}t$, the time spent in phases between ψ and $\psi+\mathrm{d}\psi$, is proportional to $\mathrm{d}\psi(1+e\cos\psi)^{-2}$, and, neglecting e^2,

$$\frac{p(\psi)}{2\pi} = \frac{1}{(1+e\cos\psi)^2} = (R/R+x)^2.$$

Introducing this value of $p(\psi)$ into (3.21) we get

$$\frac{\pi}{2}n^3V^2\nu(u, v) = \rho(x, y)p(e^2). \qquad \ldots\ (3.22)$$

We may suppose that the density of epicentres is independent of y but varies exponentially with x, or

$$\rho(x) = \rho\exp(-x/L) = \rho\exp\{v/2BL\}.$$

Since

$$u^2+\lambda^2v^2 = n^2V^2e^2,$$

$p(e^2)$ must be of the form

$$p(e^2) = f(u^2+\lambda^2v^2)$$

subject to

$$\int_0^\infty p(e^2)\mathrm{d}(e^2) = 1.$$

Equation (3.22) may be written

$$\frac{\pi}{2}n^3V^2\nu(u, v) = \rho e^{v/2BL}f(u^2+\lambda^2v^2). \quad \ldots\ldots (3.23)$$

Now

$$\nu(u) = \int_{-\infty}^{\infty}\nu(u, v)\mathrm{d}v$$

is the distribution of velocities in the u co-ordinate. Let us assume that the u velocities are observed to have an exponential distribution, or

$$\nu(u) = \frac{\nu j}{\sqrt{\pi}}e^{-j^2u^2}.$$

This implies that $f(u^2+\lambda^2v^2)$ is of form

$$\text{const} \times \exp\{-j^2(u^2+\lambda^2v^2)\}$$

and so

$$p(e^2) = j^2n^2V^2\exp(-j^2n^2V^2e^2).$$

and introducing this into (3.22) we have

$$\frac{\pi}{2}n\nu(u, v) = \rho j^2e^{v/2BL}\exp\{-j^2(u^2+\lambda^2v^2)\}. \quad \ldots\ldots (3.24)$$

Integrating (3.24) with respect to u we get

$$\begin{aligned}\frac{\sqrt{\pi}}{2}n\nu(v) &= \rho j\exp\left\{-j^2\lambda^2v^2+\frac{v}{2BL}\right\}\\ &= \rho j\exp\{-\gamma^2\lambda^2(v-v_1)^2\}\end{aligned} \quad \ldots\ldots (3.25)$$

if $v_1 = (4BLj^2\lambda^2)^{-1}$ and if $\exp(-j^2\lambda^2v_1{}^2)$ is approximately equal to unity.

Integrating (3.25) with respect to v from $-\infty$ to ∞ and remembering that $n\lambda = 2$ we have

$$\nu = \rho$$

the star density being equal to the density of the epicentres. Then

$$\frac{1}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}R} = \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}R} = -\frac{1}{L}$$

which gives

$$v_1 = \{4j^2(A-B)\}^{-1}.\frac{1}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}R}. \quad \ldots\ldots (3.26)$$

This is generally similar to equation (2.26). The treatment in the present section is restricted to v/V small and assumes that λ is constant, and has of course no equilibrium theory to establish a connexion between the potential V and the velocities. It does, however, show that a displacement term v_1 in the tangential velocities is a natural consequence of a galactic radial variation in stellar density, and that the elliptical nature of the (u, v) velocity distribution is a natural consequence of the galactic orbits.

3.3. If the density of the epicentres is not independent of y some changes on the results occur. Considering first the case where the logarithm of the density is linear, or

$$\ln \rho(x, y) = \ln \rho - \frac{1}{L}(lx + my) \qquad \ldots . (3.31)$$

where $l^2 + m^2 = 1$, we have, as in (3.24),

$$\frac{\pi}{2} n\nu(u, v) = \rho' j^2 \exp\{-j^2(u^2 + \lambda^2 v^2) - \frac{1}{2BL}(lv - mu)\} \qquad \ldots . (3.32)$$

which leads to

$$\frac{\pi}{2} n\nu(u, v) = \rho'' j^2 \exp\{-j^2(u - u_1)^2 - j^2\lambda^2(v - v_1)^2\} \qquad \ldots . (3.33)$$

where $u_1 = m/(4BLj^2)$ and $v_1 = l/(4BLj^2\lambda^2)$. The mean velocity of the stars now departs, by small amounts, from the circular velocity in both u and v co-ordinates.

Next consider the case where there is a lane among the stars, or an arm of the Galaxy rich in them. An appropriate mathematical representation of this is

$$\ln \rho(x, y) = \ln \rho - \frac{1}{L}(lx + my) \pm \frac{1}{M^2}(lx + my)^2 \qquad \ldots . (3.34)$$

the upper sign being taken for a lane among the stars and the lower sign for an arm. It is to be understood that this representation is only to apply for values of x and y small compared with the distance to the galactic centre.

We now have

$$\frac{\pi}{2} n\nu(u, v) = \rho' j^2 \exp\left\{-j^2(u^2 + \lambda^2 v^2) - \frac{1}{2BL}(lv - mu) \pm \frac{1}{4B^2M^2}(lv - mu)^2\right\} \qquad \ldots . (3.35)$$

and

$$\frac{\pi}{2} n\nu(u, v) =$$

$$= \rho'' j^2 \exp\left\{-j^2\left(1 \mp \frac{m^2}{k^2}\right)(u - u_1)^2 - j^2\lambda^2\left(1 \mp \frac{l^2}{\lambda^2 k^2}\right)(v - v_1)^2 \mp \frac{2j^2 lm}{k^2}(u - u_1)(v - v_1)\right\} \qquad \ldots . (3.36)$$

with $k^2 \equiv 4j^2B^2M^2$. The term in $(u - u_1)(v - v_1)$ produces a rotation of the axes of the velocity ellipse.

In general the quadratic

$$Au^2+2Fuv+Bv^2$$

can be transformed to its principal axes so that it becomes

$$\alpha U^2+\beta V^2$$

by a rotation θ such that

$$\tfrac{1}{2}\tan 2\theta = \frac{F}{A-B}. \qquad \ldots . (3.37)$$

Hence in equation (3.36) the principal axes of the velocity ellipse are found by a rotation θ such that

$$\tfrac{1}{2}\tan 2\theta = \pm lm/\{k^2(1-\lambda^2)\mp(m^2-l^2)\} \qquad \ldots . (3.38)$$

in the sense of decreasing longitude (since the u axis is at longitude 148° and the v axis at 58°). The ratio of the axes of the ellipse is λ_{eff} where

$$\lambda^2_{\text{eff}}(k^2\mp m^2) = \lambda^2k^2\mp l^2 \qquad \ldots . (3.39)$$

Accordingly, if any group of stars is distributed irregularly near the Sun so that there is a lane of avoidance of (the epicentres of) these stars, or an arm containing them in abundance, we expect these stars to show a velocity ellipse whose axes are tilted from the "ordinary" direction (normal and radial to the centre of the Galaxy) according to (3.38), and whose axes differ from the "ordinary" value according to (3.39). We shall later discuss whether the stars of spectral type A are distributed in this way.

4. Observed Features of Stellar Velocities

4.1. Some aspects of the observational material will now be discussed. It it is desired to work only with stars whose individual velocities are known with tolerable accuracy, it is necessary to confine oneself to stars within a comparatively small distance of the Sun, on the grounds that trigonometrical parallaxes are extremely unreliable when they are small, and, of course, the transverse velocity can only be found from the parallax and the proper motion. Accordingly, Strömberg (1946) examined the motions of 444 stars within 20 parsecs of the Sun (and with trigonometrical parallaxes greater than $0''.050$). More recently Gliese (1957) has published a catalogue of 1094 stars within 20 parsecs of the Sun, of which 597 stars (or 742 stars counting all components of multiple stars separately) have all three velocity components known. Gliese tabulates the three velocity components u directed towards longitude 148° and zero latitude: v directed towards longitude 58° and zero latitude; and w directed towards the north galactic pole. The Gliese material has been analyzed by Gliese (1956) himself and by Woolley (1958).

There are very few stars of spectral type A within 20 parsecs of the Sun (Strömberg gives 26) and for stars of this type Alexander (1958) has worked out all three velocity components by deriving a parallax from the apparent magnitude. In this work it is supposed that the spread of absolute magnitude is so small that velocities derived in this way are sufficiently reliable.

For any given catalogue of stellar velocities the mean motion relative to the Sun may be calculated without reference to any assumed velocity distribution (such as an ellipsoidal distribution). Apart from this it is necessary, in order to present the results in a compact form, to assume a definite distribution. It is now usual to assume K. SCHWARZSCHILD's ellipsoidal distribution which becomes in two dimensions

$$n(u, v) = \text{const} \times \exp -\{a(u-u_0)^2 + 2f(u-u_0)(v-v_0) + b(v-v_0)^2\} \quad \ldots . (4.11)$$

and the distribution of velocity components in *any* direction is Gaussian, as for example

$$n(u) = \int_{-\infty}^{\infty} n(u, v)\, \mathrm{d}v = \text{const} \times \exp\left\{-\left(\frac{ab-f^2}{b}\right)u^2\right\}.$$

In a Gaussian distribution $n(u) = \text{const} \times \exp(-j^2u^2)$ the dispersion σ is defined by $2j^2\sigma^2 = 1$ and so in (4.11)

$$\sigma_u^2 = \frac{1}{2}\frac{b}{ab-f^2}$$

$$\sigma_v^2 = \frac{1}{2}\frac{a}{ab-f^2}.$$

If σ_1, σ_2 are the dispersions along the principal axes of the ellipse, and the (x, y) axes are obtained by rotating the (1, 2) axes through an angle ψ,

$$\sigma_x^2 = \sigma_1^2 \cos^2\psi + \sigma_2^2 \sin^2\psi$$

$$\sigma_y^2 = \sigma_1^2 \sin^2\psi + \sigma_2^2 \cos^2\psi.$$

If it is required to find the principal axes of the velocity ellipse from the statistics of velocities in the u and v directions, we observe that in the distribution (4.1) the mean square value of $u-u_0$, or $\overline{(u-u_0)^2}$, is given by

$$\overline{(u-u_0)^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(u-u_0)^2 \exp -\{a(u-u_0)^2 + 2f(u-u_0)(v-v_0) + b(v-v_0)^2\}\mathrm{d}u\ \mathrm{d}v \Big/$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp -\left\{a(u-u_0)^2 + 2f(u-u_0)(v-v_0) + b(v-v_0)^2\right\}\mathrm{d}u\ \mathrm{d}v$$

$$= \tfrac{1}{2}b/(ab-f^2).$$

Similarly

$$\overline{(u-u_0)(v-v_0)} = -f/(ab-f^2) \text{ and } \overline{(v-v_0)^2} = a/(ab-f^2).$$

We now suppose that the sums of the observed quantities $\Sigma(u-u_0)^2$ etc., are proportional to $\overline{(u-u_0)^2}$ etc., and set

$$\Sigma(u-u_0)^2 = cb$$

$$\Sigma(u-u_0)(v-v_0) = -cf$$

$$\Sigma(v-v_0)^2 = ca$$

where c is an underdetermined constant. Then by (3.37) the principal axes are found by rotating the (u, v) axes through an angle θ given by

$$\tfrac{1}{2}\tan 2\theta = \frac{\Sigma(u-u_0)(v-v_0)}{\Sigma(u-u_0)^2-\Sigma(v-v_0)^2} \qquad \ldots . (4.12)$$

the quantities u_0 and v_0 having been found by subtracting from each u and v mean (or median) velocity relative to the Sun.

In analyzing an actual velocity distribution it is advantageous to examine the median, which is simply that velocity such that there are equal numbers in each class algebraically greater and less than the median: and the dispersion median, which is such that when differences from the algebraic median (or mean) are taken, the number of differences (without regard to sign) smaller than the dispersion median is equal to the number of differences greater than the dispersion median. In a Gaussian distribution

$$n(u) = \text{const} \times \exp(-j^2u^2)$$

we have

$$ju_m = 0{\cdot}4769$$

and

$$\sigma_u = 1{\cdot}483\, u_m.$$

The method of computing the dispersion of an actual distribution by multiplying the median difference by 1·483 has the merit that it is rapid, and is not subjective (as a graphical method would be) and finally that it does not give a high weight to very large entries, as the method of least squares does.

4.2. Table 1 shows the analysis of stellar velocities classified according to spectral type. The data for stars of type A are taken from Alexander's paper and those for the remaining types from that of Woolley.

Table 1

Analysis of Stellar Motions according to Spectral Type

Spectral type	No. of stars	Mean motion, km/sec. u_0	v_0	w_0	$\sqrt{(u_0^2+v_0^2+w_0^2)}$	Dispersion σ_u	σ_v	σ_w	σ_u^2/σ_v^2
A	475	+10	−14	−6	18·5	18	13	8	1·9
F	84	+10	−12	−6	16·4	29	16	13	3·3
G	136	+16	−20	−7	26·5	38	24	19	2·5
K	190	+15	−19	−9	26·1	37	20	16	3·4
M	292	+ 7	−16	−7	18·4	40	24	18	2·8

The relation between v_0 and σ_u^2 required by equations (2.26) and (3.36) is weak and rests on the difference between types A and F on the one hand, and types G, K and M on the other. To the values of v_0 we must add the unknown difference between the unknown circular velocity and the solar motion. If

for early type stars $v_0 = -13$ km/sec, $\sigma_u^2 = 529$ (km/sec)^2
for late type stars $v_0 = -18$ km/sec, $\sigma_u^2 = 1521$ (km/sec)^2

and if $v_\odot$ is the excess of the solar motion over the circular motion, then for each class $v_1 = v_0 + v_\odot$. By (3.26)

$$v_0 + v_\odot = v_1 = -\sigma_u^2/\{2L(A-B)\}$$

and A–B may be taken as 33 km/sec/kpc. We now have simultaneous equations for L and $v_\odot$ which are satisfied by

$$L = 3 \text{ kpc} \qquad v_\odot = +10 \text{ km/sec}$$

the figures obtained being reasonable except that L is rather large to judge by SCHMIDT's model (1956) of the Galaxy from which one would estimate $L \sim 1$ kpc. Since the value of v_0 for all GLIESE stars is -17 km/sec. according to WOOLLEY, the Sun leads the GLIESE stars by 17 km/sec. and with $v_\odot = +10$ km/sec the mean of the GLIESE stars trails behind the circular velocity by 7 km/sec.

It is, however, clear from Table 1 that the solution is not of great weight, since σ_u is quite different for the A and F stars although v_0 is nearly the same for the two spectral types. If we disregard the F stars and also suppose the $\sigma_u{}^2$ for the late type stars is over-estimated (on account of a selection effect that will be briefly discussed later) and set $\sigma_u{}^2 = \frac{1}{2} \times 1521$ for the late type stars, we get $v_\odot = +11$ km/sec and $L = 1{\cdot}6$ kpc. These modifications show the amount of uncertainty in the calculations.

4.3. The selection effect referred to is due to the fact that faint stars are usually put on parallax programmes because they show a large proper motion. It must be supposed that there are a number of faint late type stars (dwarf stars of spectral type K and M) within 20 parsecs of the Sun which have never been distinguished from giants simply because they have a small transverse velocity and have not been observed for trigonometric parallax. Accordingly the velocity dispersion of the stars known to be within 20 parsecs of the Sun is higher than the dispersion of all such stars. The point is a well-known one and may be tested by comparing the transverse velocities with the radial velocities of the same stars. For an isotropic and Gaussian velocity distribution the mean transverse velocity is $\sqrt{2}$ times the mean radial velocity and for an ellipsoidal distribution approximating to the observed distribution this ratio is $\sqrt{(2{\cdot}156)}$, whereas the observed ratio of tangential to radial velocities is $\sqrt{(5{\cdot}2)}$, for M stars (WOOLLEY, 1958). This indicates that the velocity dispersion is over-estimated by perhaps 100 per cent.

4.4. Instead of breaking up the stars into groups according to their spectral type we may group them according to the eccentricities of their orbits. An immediate reason for doing this is the consideration that the theory of section 3 only refers to orbits with small eccentricity. Table 2 shows various data found in this way.

The point examined in this section is the tilt of the axis of the elliptic velocity distribution. By (4.12) this has the same sign as the sum of products $\Sigma(u-u_0)(v-v_0)$. As is shown in Table 2 the tilt is small in the groups with smallest eccentricity and indeed is not of constant sign among them, from which it may be deduced that in the case of stars whose orbits have small eccentricity the tilt is negligible and that the stars are uniformly or monotonically distributed without any arm or lane among them, such as discussed in section (3.3); or, in the notation of that section, $M = \infty$.

However, the A-type stars do not behave in this way. This can be seen in a diagram given by STRÖMBERG (1946). STRÖMBERG shows tilted ellipses for stars of all

spectral types; he does not exclude orbits of large eccentricity and does therefore include some fast stars which make large contributions to the product sum $\Sigma(u-u_0)(v-v_0)$. In Strömberg's diagram the tilt is much greater for stars of spectral type A than for other spectral types. This is confirmed by Alexander, who finds a large tilt and moreover finds that this tilt does not disappear when orbits of large eccentricity are excluded.

It seems possible to conclude that the A type stars are distributed with their epicentres in an arm or lane in the sense of section (3.3), but that stars of later spectral type are much more evenly distributed. In the later types the distribution of velocities is (3.24) or (3.33); and in either case we have $\sigma_u{}^2/\sigma_v{}^2 = \lambda^2$, and λ^2 is

Table 2

Classification of orbits of 743 stars within 20 parsecs of the Sun according to eccentricity

Class	a	b_1	b_2	c	d	e	e$^+$
Limits of $S = 2\text{BeR}$	0 to 14	14 to 21	21 to 28	28 to 42	42 to 56	56 to 70	>70
No. of stars	159	151	137	155	70	30	41
Average mass, $\odot = 1$	0·94	0·82	0·79	0·72	0·64	0·75	0·66
Mean $v-v_0$, unweighted for stellar mass.	+1·3	+2·5	+4·8	−3·2	−13·9	−63·8	−41·9
Mean $v-v_0$, weighted	+1·7	+2·9	+6·7	−2·7	−15·0	−20·6	−76·1
Tilt θ	+4°·2	+0°·6	−6°·0	−10°·8	−7°·3	+1°·7	

Note.—S is computed from equation (3·181) by $S^2 = 1/\lambda^2(u-u_0)^2+(v-v_0)^2$ with $u_0 = +11$ km/sec and $v_0 = -17$ km/sec. The angle θ is computed from equation (4·12) with the same values of u_0 and v_0 except class d, $v_0 = -31$ km/sec, and classes e and e^+, $v_0 = -62$ km/sec.

related to the Oort constants A and B by $\lambda^2 = (B-A)/B$. From Table 1 the value of λ^2 is about 3 for late-type stars. If the Oort constant A is taken to be well determined from the radial velocities of distant objects and to have the value +20 km/sec/kpc, then the constant B is −10 km/sec/kpc. On the other hand, for the A stars the value for $\lambda^2{}_{\text{eff}}$ is approximately 2 (Table 1). If we set $\lambda^2 = 3$ and $\lambda^2{}_{\text{eff}} = 2$ in equation (3.39) we get

$$k^2 = \pm(l^2-2m^2) \qquad \ldots\ (4.41)$$

According to Alexander the axis of the ellipse for A stars is tilted about 30° from the direction of the centre of the Galaxy, in the direction of increasing longitude. Then

$$\tfrac{1}{2}\tan 2\theta = -\frac{\sqrt{3}}{2}.$$

From (3.37) with $\lambda^2 = 3$ we get

$$l^2 - \frac{2}{\sqrt{3}}lm - 3m^2 = 0$$

and $l/m = (\sqrt{10}+1)/\sqrt{3}$ (solution (a)) or $l/m = -(\sqrt{10}-1)/\sqrt{3}$ (solution (b)). Since k is real, solution (a) requires the upper sign in equations (3.34) and (4.41), and the solution refers to a lane. Solution (b) refers to an arm. In solution (a) the numerical value of l/m is 2·403 so that the lane is normal to the direction inclined at an angle $\cot^{-1}2{\cdot}403$ or 23° to the u-axis turning towards the v-axis: that is to say,

normal to longitude $148° - 23°$ or $125°$. In solution (b) the numerical value of l/m is $-1{\cdot}248$ so that the arm is normal to $187°$.

The lane runs in the directions $35°$ and $215°$ longitude. By (4.41) the value of k is $0{\cdot}74$. Since $k^2 = 4j^2B^2M^2 = 2B^2M^2/{\sigma_u}^2$ this value of k leads to a value of M. Taking $B = -10$ km/sec/kpc and $\sigma_u = 18$ km/sec for the A stars, we find $M = 0{\cdot}94$ kiloparsecs.

The representation of the actual distribution of stars by an arm in the Galaxy or by a lane is, of course, an over-simplification. The distribution of hydrogen as shown by 21 cm radiation does suggest a lane, generally in the direction of $35°$ longitude, (VAN DE HULST, MULLER and OORT (1954)), but the distribution of early type stars given by MCCUSKEY (1956) suggests rather an arm in the general direction of $97°$ longitude. It seems quite likely that the tilt of the velocity ellipse of the A stars is in fact connected with irregularity of the space distribution of these stars.

5. PERMANENCE OF VERY OPEN CLUSTERS

A question of interest is the permanence, or otherwise, of open clusters—that is to say, of groups of stars travelling together at such distances from one another that they are not held together by mutual gravitational attraction. The idea entertained formerly (for example by EDDINGTON when he wrote *Stellar Movements and the Structure of the Universe*, 1914), was that the stars moved with parallel velocities. The requirement of parallelism was severe. For since 1 km/sec equals 1 parsec in 10^6 years, and since the groups were supposed to be 10^{10} years old, a group seen to occupy a volume of diameter less than 10 parsecs must possess velocities with differences of order 10^{-3} kilometres per second. We no longer suppose that all stars are 10^{10} years old, but there may well be groups at least 10^8 years old. Further, stars in quite different parts of the sky are supposed to be travelling with common velocity. The group of stars apparently travelling with the Hyades has been discussed recently by EGGEN (1957) and the order of magnitude of the diameter of the group discussed is at least 100 parsecs. Over such distances the requirement of strict parallelism becomes modified. A star moving with the motion (3.11) repeats its distance from the centre of the Galaxy when $n\theta$ has increased by 2π. Calling the time to perform this the period P_1, we have

$$P_1 = \frac{h^3}{{\mu_1}^2}\int_0^{2\pi}\frac{\mathrm{d}\psi}{1+e\cos\psi} = \frac{2\pi h^3}{{\mu_1}^2}(1-e^2)^{-\frac{1}{2}}$$

or, neglecting e^2 and using $h = \varpi_1\Theta_c$

$$P_1 = 2\pi\varpi_1/\Theta_c$$

the period of the epicentre of the star's orbit. Considering orbits with different distances of their epicentres from the galactic centre,

$$\frac{\mathrm{d}P_1}{P_1} = \frac{\mathrm{d}\varpi_1}{\varpi_1}\left(1 - \frac{\varpi_1}{\Theta_c}\frac{\mathrm{d}\Theta_c}{\mathrm{d}\varpi_1}\right) = \frac{\mathrm{d}\varpi_1}{\varpi_1}\left(\frac{2A}{A-B}\right).$$

The value of $(2A/(A-B))$ is about $4/3$. Two stars whose epicentres are separate in distance from the galactic centre by one per cent will have periods differing by $4/3$

per cent, and will separate in distance transverse to the galactic centre by $2\pi \times 4/3$ per cent or about 8 per cent of the distance to the galactic centre, in each revolution round the galactic centre: or, by about 800 parsecs in 10^8 years. Hence for stars in an open cluster to stay together within a distance of 100 parsecs in a life of 10^8 years the *periods* must be equal within one eighth of one per cent, and so on. The condition of *parallel velocity* postulated in the absence of galactic forces and valid over small interstellar distances must be generalized to a condition of *equal galactic period* with rather severe tolerances.

Consider an origin at a distance R_1 from the galactic centre moving with velocities u_1 and v_1 relative to the circular motion V_1 at R_1. If a star has co-ordinates (x, y) and velocities (u, v) relative to this origin, the star has a (tangential) velocity

$$v - v_1 - x\frac{\mathrm{d}V}{\mathrm{d}R}$$

relative to the circular velocity at its own place, and its epicentre is, by (3.17), located a distance

$$-\frac{1}{2B}\left(v - v_1 - x\frac{\mathrm{d}V}{\mathrm{d}R}\right)$$

inside its present position. The x co-ordinate of the epicentre relative to the origin at R_1 is therefore

$$x + \frac{1}{2B}\left(v - v_1 - x\frac{\mathrm{d}V}{\mathrm{d}R}\right).$$

The condition that a group of stars should have equal galactic periods is the condition that their epicentres should all have the same distance from the galactic centre, or

$$x\left(1 + \frac{1}{2B}\frac{\mathrm{d}V}{\mathrm{d}R}\right) - \frac{v}{2B}$$

constant for the group. The condition of equal period is accordingly

$$\frac{\mathrm{d}v}{\mathrm{d}x} = 2B + \frac{\mathrm{d}V}{\mathrm{d}R} = B - A. \qquad \ldots\ldots (5.1)$$

In the ordinary formula for galactic rotation (in stars all proceeding with the circular velocity) we have

$$\frac{\mathrm{d}V}{\mathrm{d}R} = -(A + B).$$

A group of stars with equal galactic period accordingly shows the phenomenon of galactic rotation with the coefficient $-(A-B)$ instead of $-(A+B)$. Since A is positive and B negative, the group rotation is greater than galactic rotation perhaps by three to one (the value of B being uncertain). If $A - B = 30$ km/sec/kpc the velocity gradient is 3 km/sec per 100 parsecs, a quantity hard to determine but perhaps just observable in actual very open clusters.

If stars in open clusters do stay together by reason of their possessing nearly equal galactic periods, the stability of the cluster is due to the initial motions of the stars when first formed being such as to lead to equal periods; that is to say that the stars possessed these equal periods if and when they formed by condensation out of a large interstellar gas cloud. If the original gas cloud was large and withstood the same galactic shearing forces as would break up an open cluster, it must have done so by possessing the appropriate rotation (5.1). One may either suppose that gas clouds do or do not possess this rotation by chance, and if they happen to do so, they form stable clusters: or else that the cloud is somehow held together (for example by magnetic fields—a suggestion I owe to Dr. PAGEL) and that the interaction between cohesive and gravitational forces compels the cloud to rotate in such a manner that its various parts have equal galactic periods.

6. CONNECTION BETWEEN ORBITS AND COLOUR-LUMINOSITY ARRAYS

By equation (3.19) the distance from the Sun to the centre of the Galaxy being R, the closest approach made by any star to the centre of the galaxy is

$$R+\frac{1}{2B}\left[\sqrt{\left(\frac{u^2}{\lambda^2}+v^2\right)}-v\right] \qquad \ldots . (6.1)$$

$2B$ being a negative quantity. The theory which gives this result neglects the squares of the eccentricities of the orbits and is therefore unreliable when the star penetrates a quarter of the way to the galactic centre or more, but nevertheless classification of perigalactic distances with the help of this equation is some guide to the approaches made by various groups of stars to the galactic centre, even if the division points given by (6.1) are inexact.

Accordingly WOOLLEY and EGGEN (1958) classified the stars in the solar neighbourhood by computing the velocity

$$\mathbf{v} = -v+\sqrt{\left(\frac{u^2}{\lambda^2}+v^2\right)}.$$

The stars were then classified as follows:

Class A	$\mathbf{v}=$	0 to	9	km/sec.
B		10 ,,	29	,,
C		30 ,,	59	,,
D		60 ,,	99	,,
E		100 ,,	129	,,
E$^+$		> 130		,,

If one adopts $-2B = 20$ km/sec/kpc and $R = 10$ kpc, the division point between classes D and E is halfway in to the galactic centre from the Sun: but as has been said this is inexact, although the order A, B, C, etc., does represent increasing approach to the galactic centre.

Having classified the stars in this way WOOLLEY and EGGEN constructed colour luminosity arrays for the six classes. These are reproduced on pp. 272–274 of this volume 3 of *Vistas in Astronomy*. The arrays show a gradual change from that of class A,

which resembles that of the Pleiades, to that of class E^+, which resembles that of the open cluster M 67. In class A there is a main sequence extending to stars of spectral type A, and in class E^+ there are no stars of spectral type earlier than G, the giant stars being red.

In terms of current evolutionary theory the Pleiades is a new cluster, while M 67 is an old cluster in which the giant stars have left the main sequence to become red giant stars. Accepting this interpretation the stars with more or less circular orbits in the neighbourhood of the Sun (and those whose orbits lie mainly outside the Sun), that is the stars of class A, are of comparatively recent formation, but those which are now in the solar neighbourhood but have come out from the galactic centre, the stars of classes E and E^+ are considerably older. This state of affairs agrees well with Baade's observation that the giant stars in the centre of the Andromeda nebula are red, and the interpretation which is put upon this, that the stellar population at the centre of a galaxy is old and the population in the outer parts new—star formation, in fact, starting at the centre of a galaxy and proceeding outwards. The old stars in the solar neighbourhood are distinguished by their high velocities (specifically by high values of the derived velocity **v**) and are interlopers in the solar neighbourhood coming from the inner parts of the Galaxy.

7. Equipartition of Energy

We may refer here briefly to the suggestion of equipartition of energy conveyed in Table 1, which is simply that σ^2 increases from spectral type A to the later spectral types, while the masses of the A stars are greater than the masses of the K and M stars, if these are dwarfs, which is true of the great majority of stars in the solar neighbourhood. We have, however, seen that selection by proper motion increases the value of σ^2 for the late type stars: also it is not clear that velocity *differences* from the circular velocity should be used in this way. Velocities in the direction perpendicular to the galactic plane are free from the latter objection. Among the nearby stars in Gliese's list there is considerable variation in mass (deduced from the absolute magnitude by the mass-luminosity law) and when the w-velocity dispersions are classified according to stellar mass there is no significant variation (Woolley, 1958, Table VI). The attempt to establish equipartition of energy by classification according to spectral type is probably misleading, as the A stars behave in a peculiar way, as is shown by the peculiar tilt of their velocity ellipse. Moreover, stars of spectral type A are unrepresented in the fast star classes C to E^+ which come from the centre of the Galaxy. Massive stars in these classes have become red.

It is worth referring to this topic mainly because the idea of equipartition has been quite firmly held in the past; cf. Bok (1946). Contemporary estimates of the time of relaxation of velocities due to stellar encounters in the solar neighbourhood are of order 10^{12} years, and therefore much older than the ages of the stars (not more than 10^{10} years), so that it is now generally held that there has not been time to set up equipartition of energy in the neighbourhood of the Sun.

It may, however, be unwise to regard this question as entirely closed, and Ogorodnikov (1957) has recently argued that the time of relaxation of galaxies is only of order 10^7 or 10^8 years. (The classical value of 10^{12} years or more is deduced from star-to-star encounters, and the shorter time of relaxation can be obtained if it

is assumed that 10 per cent of the matter in galaxies is accumulated into condensations of 10^5 or 10^6 solar masses.) OGORODNIKOV argues that the SCHWARZCHILD ellipsoidal velocity distribution and the observed uniform rotation of galaxies require the short relaxation-time scale. The former argument is not at all supported by section 2 and 3 of this article.

Nevertheless the fact that "older" stellar populations exhibit greater velocity dispersions than "younger" stellar populations is probably an argument for disturbance to stellar velocities more vigorous than star-to-star encounters. Possibly some reconciliation of various points of view can be effected by recognizing that there is not a single time of relaxation for all aspects of disturbance to stellar velocities.

REFERENCES

ALEXANDER, J. B.	1958	*M.N.* **118,** 161.
BOK, B. J.	1946	*M.N.* **106,** 66.
EGGEN, O. J.	1957	*Observatory* **77,** 229.
GLIESE, W.	1956	*Z. Astrophys.* **39,** 1.
	1957	*Heidelberg Astron. Rechen-Institut*, Mitt. Serie A, No. 8.
MCCUSKEY, S. W.	1956	*Ap. J.* **123,** 458.
OGORODNIKOV, K. F.	1957	*Astron. J. USSR* **34,** 809.
SCHMIDT, M.	1956	*Bull. Astron. Inst. Netherlands*, **13,** No. 468.
SMART, W. M.	1938	*Stellar Dynamics*, Cambridge University Press.
STRÖMBERG, G.	1946	*Ap. J.* **104,** 12.
VAN DE HULST, H. C., MULLER, C. A. and OORT, J. H.	1954	*Bull. Astron. Inst. Netherlands*, **12,** No. 452, p. 117.
WOOLLEY, R. V. D. R.	1958	*M.N.* **118,** 45 and 198.
WOOLLEY, R. V. D. R. and EGGEN, O. J.	1958	*M.N.* **118,** 57.

The Rediscovery of Neptune

R. A. Lyttleton, F.R.S.

St. John's College, Cambridge, England

Summary

The methods used by Adams and Le Verrier for the prediction of the position of Neptune led to extremely extensive calculations, and the question still remains open whether any much simpler approach is possible that could have led to comparably accurate prediction with much reduced effort of calculation. Such a method is presented. It turns out that the time of conjunction with Uranus can be derived with very little calculation using only elementary considerations applied solely to the (observed) discrepancies in heliocentric longitude. The instant so found is only about six months later than the actual value. On the basis of a circular orbit for the unknown planet, its position at discovery (25 years later than conjunction) even using Bode's law for the size of orbit is then only some 13° away from the actual position, and in fact within the zone that Challis began to search, and so probably sufficient for ultimate discovery. However, the distance appropriate to a best-fit of the observations can also be found, and this makes a considerable improvement in the predicted position. By suitably combining the equations of condition, the number of unknowns associated with the necessary correction of the orbit of Uranus can be reduced from four to two, and with any assumed radius for the (circular) Neptune orbit only the mass of the planet is unknown. Thus the number of unknowns can be reduced to three (for each assumed mean distance) compared with eight in the original methods. Even with the extra work of calculation of the coefficients in the perturbations for a sufficient number of assumed mean distances, there results an enormous reduction in the amount of calculation. For the demonstration of the method, further economy of arithmetic has been achieved by making use only of a part of the observational material that was actually available at the time (the so-called "modern" observations). Nevertheless the procedure succeeds in predicting a position of Neptune (in 1846) with the same accuracy (1°) as Le Verrier and more accurately than Adams ($2\frac{1}{2}$°).

1. Introduction

The planet Uranus was discovered in 1781, and Neptune in 1846 as a result of studies of the discrepancies in the motion of the former unexplained by the action of other already known planets, here Jupiter and Saturn. The hypothesis of a further planet superior to Uranus was an obvious possibility, but the steps needed to proceed from the bare hypothesis to the actual discovery were apparently beyond the resource and courage of most astronomers of the day. Adams and Le Verrier alone carried through the requisite mathematical labours that seemed to them to be involved, though Bessel was also preparing to attack the problem in the months preceding his death. However, for the purposes of the present discussion, it is important to emphasize at the outset that the success achieved by both Adams and Le Verrier was only partial. Each of them had aimed not only at discovery of the planet but also at the calculation, from the known residuals of the motion of Uranus, of the whole set of orbital elements of the hypothetical planet and its mass, assuming only that it moved in a path coplanar with that of Uranus and at a mean distance given approximately by Bode's law. Had these latter elements been satisfactorily determined, the position of the planet at any time would follow, and its telescopic discovery made possible. But in fact the orbital elements they found could not be regarded even as approximations to the true values, later ascertained. With sufficient accuracy to demonstrate the differences, their values were as follows:

Table 1

	Adams		Neptune	Le Verrier
	I	II		
Mean distance	38·4	37·2	30·07	36·2
Longitude of perihelion	316°	299°	44°	284°
Eccentricity	0·161	0·121	0·0086	0·108
Sun/Mass	6040	6660	19,300	9350

Nevertheless, these values enabled the position of the planet in the year of discovery 1846 to be predicted to within about 1° by means of the elements found by Le Verrier, and to within about $2\frac{1}{2}$° by those given by Adams. The ultimate cause of these erroneous elements lay of course in the adoption of Bode's law for an empirical indication of the mean distance; unfortunately for the investigators, although the law would have closely predicted the mean distance of Uranus before its discovery, it entirely fails for Neptune. Thus the achievements of Adams and Le Verrier were in fact limited simply to that of discovery *at the time*, and did not include the prior evaluation of the orbital elements.

There can be little doubt that both Adams and Le Verrier must have begun by thinking in terms of a circular orbit, just as they assumed throughout an orbit coplanar with that of Uranus, but it would soon have become clear to them that such an orbit *with radius given by Bode's law* could not hope to satisfy the observations. The introduction of an eccentricity, and with it a longitude of perihelion, both to be determined subsequently, became inevitable, and led them as it did into extremely extensive calculations. Further complexity enters from the fact that the elliptic elements already derived for Uranus, after removal of the known effects of Jupiter and Saturn, were necessarily in error by amounts depending principally on the mass of the unknown perturbing planet, and so it becomes necessary to include as further unknowns corrections to the four elements of the (coplanar) orbit of Uranus. This leads to equations involving eight unknowns, even after the assumption at the outset of a definite value for the mean distance (of Neptune).

The resulting calculations were applied to observations not only from 1781 to 1840 (in the work of Le Verrier to 1845), but also to the so-called *ancient* observations of Uranus dating back to 1690, which refer to unnoticed recorded observations of it prior to discovery and retrieved *afterwards*, and with this long period of observations available the resulting numerical work was extremely lengthy.

The successful outcome, in the matter of actually finding the planet, naturally aroused great enthusiasm, and understandably deflected attention from the interesting technical question raised by the fact that the actual orbit of the planet, soon after computed, differed so much from the predicted ones. In describing the work as representing the very zenith of mathematical skill, there was possibly some element of pardonable exaggeration in the atmosphere of great excitement that such a discovery at such a time must have created. But even in the centenary celebrations of the event in 1946 there was some repetition of the lavish praises of a century earlier. It was at this time that the present investigation had its inception. In discussion with J. E. Littlewood,[1] the author ventured the opinion that it might well

[1] For Professor Littlewood's own account of the matter see his *A Mathematician's Miscellany*, p. 117 *et seq.*, Methuen, 1953.

be possible to have discovered the planet by much more direct and elementary methods not involving such excessive calculations, and found that this same view had been held by J. E. LITTLEWOOD himself for many years. In virtue of the erroneous orbits found by ADAMS and LE VERRIER, we agreed that the modern form of the problem could be put as follows:

> *What is the simplest theoretical approach and minimum amount of calculation that could have led to the discovery of the planet?*

The answering of this question is the object of the present paper. We shall take the requirements of this question as our guide throughout, and so from the outset abandon any idea of attempting to calculate the orbit of Neptune in detail. Instead, the sole object will be the prediction of its position on 23 September 1846, the date of discovery. Also, wherever possible we shall endeavour to reduce the amount of arithmetic, using less observational material than was actually available, using fewer equations, and so on. The various economies will become apparent as we proceed.

2. THE OBSERVATIONAL MATERIAL

The principal quantity actually showing the discrepancies produced by the unknown planet is the heliocentric longitude of Uranus, and this of course has to be deduced from the actual observations of the planet. The basic material that will be used consists of the differences in heliocentric longitude as between that found from observations and that given by the ellipse of closest fit for the period prior to 1840 (that is after the removal of the effects of Jupiter and Saturn). This material has been tabulated by ADAMS[2] at intervals of three years at the beginning of each particular year, as follows:

Table 2

Discrepancies in heliocentric longitude, $\delta(t)$, for Uranus (Observed—Calculated)

1780	3″46	1801	22″21	1822	20″97
83	8″45	04	24″16	25	18″16
86	12″36	07	22″07	28	10″82
89	19″02	10	23″16	31	− 3″98
92	18″70	13	22″00	34	−20″80
95	21″38	16	22″88	37	−42″66
98	20″95	19	20″69	40	−66″64

The present work has not used any other data beyond these, making in fact no use of the observations at intervening times, nor of the ancient observations which in fact extend over nearly a century prior to discovery. While this may result in some loss of accuracy, it involves considerable economy of arithmetic, and thereby is in accord with the spirit of the enquiry. Also we are here concerned primarily with method rather than with achieving the utmost possible accuracy of prediction.

[2] J. C. ADAMS, Collected Works, p. 11, Cambridge 1896.

This observational material when plotted on a graph (Fig. 1) gives a series of points through which it is clearly impossible to draw accurately any reasonably smooth curve, and it is plain from any such attempt that the above figures must be subject to errors of the order of 0″.5 (omitting what appears to be an unduly large error of about 3″.0 in the value for 1789).[3] In the treatment we adopt, a smooth series of values is desirable for two reasons. First, the procedure to be explained for determining the time of conjunction in longitude of Uranus and Neptune depends on second differences of the discrepancies in longitude, and these tend to magnify

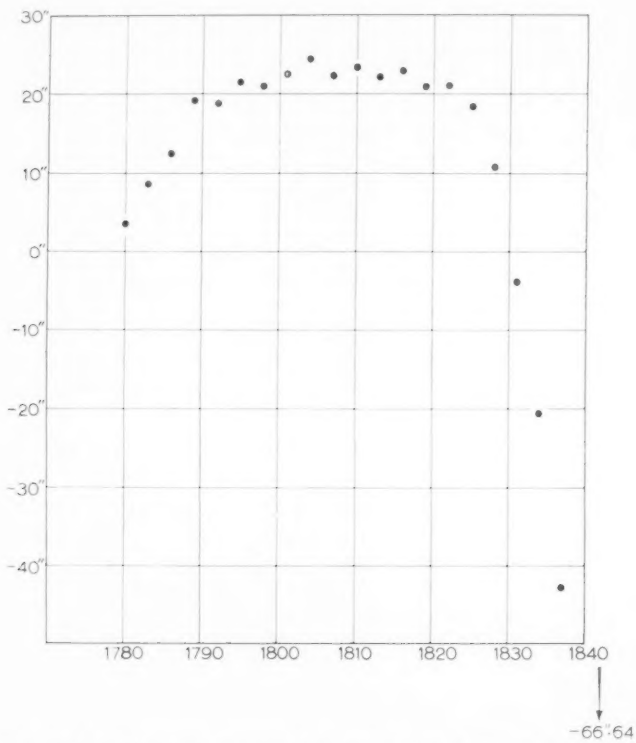


FIG. 1. Discrepancies in heliocentric longitude of Uranus (1780–1840) at 3-year intervals in seconds of arc (ADAMS, Coll. Works, p. 11). These are the quantities $\delta(t)$ of Table 2.

random errors; and second, the corresponding values (to Table 2) for intermediate years, viz. 1781, 1782, 1784, etc., are later required and are obtained by simple interpolation between the above values when these have been smoothed. Now inspection of the points of Fig. 1 makes clear that although it is possible to draw a smooth curve to cover the observations from 1780 nearly to 1800 (again with the exception of 1789 which must be too large a value on any basis) and again from about 1820 on to 1840, it is difficult to draw any unique such curve to represent the intermediate points from about 1800 to 1820. In fact of course the accurate values would lie on a smooth curve the form of whose equation could be determined dynamically only were everything already known about the system; in other words knowledge of

[3] This value could be an error of copying for 16″.02.

even more than the things we seek to find. To make a least-squares fit, some analytical form must be assumed, and as any such form cannot be adequate, the actual course we decided to adopt is to represent the observations between 1801 and 1819 by a straight line parallel to the abscissa at such a level (in fact at 22″.75) given by the mean height (ordinate) of all the points in this interval. This appears to provide the simplest way of getting over the difficulty of drawing the "proper" curve through the points and of avoiding any tendency to shift the observations in any systematic way that might introduce bias. By this means the perfectly definite curve of Fig. 2 was arrived at as representing the observations. The values for the years of Table 2 are then corrected to the curve values, while those required for intervening years can now be taken directly from the curve quite unambiguously as required. In this way we arrive at the following values which will be used as the basis of the calculations of this paper:

Table 3

Smoothed values of $\delta(t)$

1780	3″.46	1796	21″.45	1825	18″.16
81	5″.0	97	21″.8	26	17″.0
82	7″.0	98	22″.00	27	14″.6
83	8″.45	99	22″.25	28	10″.82
84	9″.8	1800	22″.50	29	7″.8
85	11″.5	1801		30	2″.0
86	12″.80	to	22″.75	31	− 3″.98
87	14″.1	1816		32	− 9″.6
88	15″.25	1817	22″.76	33	−15″.2
89	16″.40	18	22″.78	34	−20″.80
90	17″.45	19	22″.80	35	−28″.08
91	18″.25	20	22″.4	36	−35″.36
92	19″.00	21	20″.87	37	−42″.66
93	19″.9	22	20″.97	38	−50″.66
94	20″.6	23	20″.3	39	−58″.65
95	21″.0	24	19″.6	1840	−66″.64

The method we shall use to "predict" the position of Neptune for the actual day of discovery falls into two separate parts. The first stage of the calculation is concerned solely with the determination of the time of heliocentric conjunction of the two planets. This having been found (in fact within about 6 months of the actual instant), it turns out that a considerable range of assumed mean distance for Neptune would predict the position for discovery within, say, $\pm 15°$ of longitude without further calculation. The second part of the investigation is then concerned with the question of finding the best estimate of mean distance on the assumption always that the unknown planet is moving in a circular orbit. Such an assumption is quite as reliable a "law" for planetary orbits as is BODE'S, and as already remarked might equally well have been adopted by ADAMS and LE VERRIER had they not been concerned with the difficulty (see later p. 37) brought about by the manner in which the mass and mean distance are together involved in the dynamical theory. In the line of attack pursued in the second stage of the present work, however, where a succession of assumed values of the mean distance is adopted, the mass enters only linearly, its coefficient depending intricately on a/a', the ratio of mean distances, and it can be solved for by least squares in a straightforward way.

Accordingly we shall at this later stage work with the assumption of a strictly circular orbit for the unknown planet.

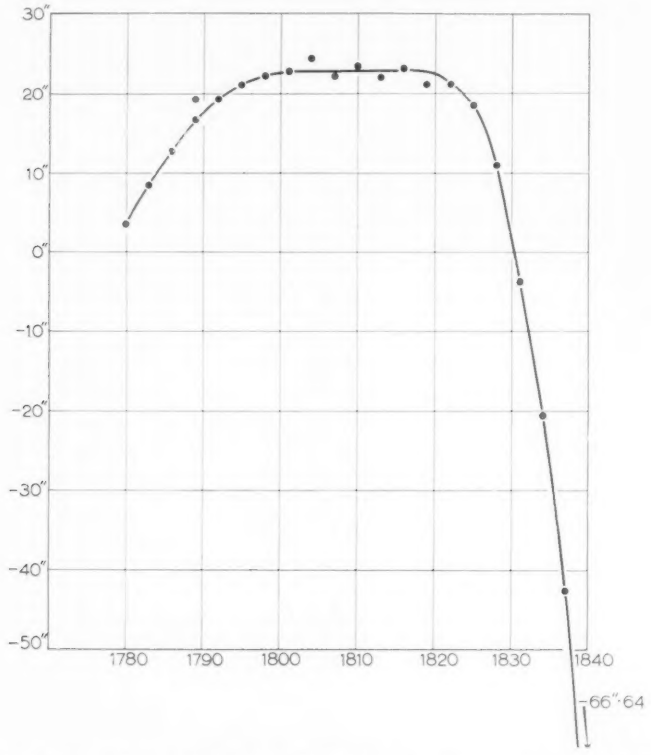


FIG. 2. Smoothed values of the observations 1780–1840 (straight line from 1801 to 1816, and smooth curves at each end of range). This curve gives the values of Table 3.

3. NOTATION

In accordance with the standard notation of planetary theory we shall use unaccented letters m, a, n, e, ϵ, ϖ for the mass and elements of the inner planet, viz. Uranus, and accented letters m', a', n' for the required elements of the unknown planet. The Sun's mass is taken as unity. The time will usually be denoted by t.

The assumption made by both ADAMS and LE VERRIER of coplanar orbits in the plane of the ecliptic is adopted throughout of course.

4. DETERMINATION OF THE TIME OF HELIOCENTRIC CONJUNCTION

The most direct dynamical property that could conceivably lead to a determination of the instant of conjunction is clearly the variation of angular momentum of Uranus, which at any stage is defined in the usual notation by $r^2\dot{v} = h$. For a considerable time before conjunction Neptune will be ahead of Uranus in longitude and so dh/dt will be positive, whereas after conjunction it will be negative. Immediately on either side of conjunction, however, the rate of change of h will be small since the pull of Neptune will be almost along the radius vector of Uranus. Thus the curve of h

against t will be very flat-topped at and near conjunction anyway. Nevertheless, if h were known with sufficient accuracy over a range containing the time of conjunction, its derivative would vanish at the instant of conjunction. The idea fails in practice for the present case, however, because although v is well determined, the radius vector r is not capable of determination with anything approaching the same accuracy, and the resulting plot of $r^2\dot{v}$ against time does not produce a curve sufficiently smooth and regular to allow the instant of vanishing gradient to be read off at all precisely. But it happens that a far more accurate alternative procedure based entirely on considerations of the longitude v of the planet is possible.

To explain this, consider a planet m moving in an accurate newtonian elliptic orbit of *small* eccentricity e. Then its heliocentric longitude v at any time t is given by the well-known formula

$$v = nt + \epsilon + 2e \sin(nt + \epsilon - \varpi), \qquad \ldots . (1)$$

wherein we agree to neglect all terms in e^2 and higher powers of e. (For Uranus the value of e is 0·047). If now for any reason values of n, ϵ, e, and ϖ were adopted that differed from these by small amounts depending on the mass m' of a perturbing planet and v calculated by means of these slightly different elements, the resulting error in longitude would be given by

$$\begin{aligned} \Delta v = {} & t\Delta n + \Delta\epsilon + 2\Delta e \sin(nt + \epsilon - \varpi) \\ & + 2e(\Delta\epsilon - \Delta\tilde{\omega}) \cos(nt + \epsilon - \tilde{\omega}) + 2et\Delta n \cos(nt + \epsilon -)\varpi. \end{aligned} \qquad \ldots . (2)$$

Each of the first four terms on the right, including that in $e(\Delta\epsilon - \Delta)$, $\varpi$is of order m' if the differences of the elements are regarded as due to perturbations by m'. But the last term in (2) is of order em', since Δn is of order m'. Accordingly this term is small compared with the others and can be neglected. It is in fact this feature of the expression for Δv that permits the time of conjunction to be found with reasonable accuracy by elementary methods, as will appear below. Hence, on the basis that we neglect all terms $O(em')$, we have

$$\Delta v = m'(a + bt + c \cos nt + d \sin nt) \qquad \ldots . (3)$$

where a, b, c, and d are constants simply related to the elliptic elements appearing in (2). Also, it may be noted that any approximation to the value of the mean motion n that differs from it only by an amount of order m' can equally well be used in (3) without affecting the form for Δv, since the effect of any such change will be of order m'^2 on Δv.

Consider now an elliptic orbit E fitted to the observations of Uranus (m) *after* the removal of all known dynamical perturbations (that is, after removal of the effects of Jupiter and Saturn, which accordingly need not hereafter be referred to). These observations will *not* of course be capable of proper representation by such an ellipse since they contain the effects of the unknown Neptune (m'). Let us write then

$v_E(t)$ = longitude of m in the ellipse E best fitting the observations,

and

$v(t)$ = the observed longitude of Uranus, m.

Then the discrepancy, δ say, in longitude (Tables 2 and 3) at any time may be written

$$\delta(t) = v(t) - v_E(t). \qquad \ldots . (4)$$

Now let v_0 denote the longitude that Uranus would have in its instantaneous elliptic orbit, E_0 say, corresponding to the (unknown) time of conjunction t_0, say, which it is our prime object to determine. The ellipse E_0 is the path that Uranus would describe from t_0 onwards (and backwards) in time if the action of m' were to cease at the instant t_0. The orbit E_0 is of course not known, but it will clearly differ from E and from the actual path of m by terms of order m' and higher.

Also let $w(t)$ denote the perturbation in the longitude v of Uranus produced by Neptune *since time* t_0. (The value of $t-t_0$ can be both positive and negative of course.)

To clarify these definitions by setting them out, we have then at any time t

$v(t)$ = *observed* longitude of Uranus,
$v_E(t)$ = *calculated* longitude of Uranus from the *known* best-fitting ellipse E,
v_0 = the (unknown) longitude of Uranus in the instantaneous ellipse E_0 (unknown),

and by definition

$$w = v - v_0.$$

Hence we have from (4) that

$$\begin{aligned} \delta(t) &= v_0 - v_E + (v - v_0) \\ &= (v_0 - v_E) + w. \end{aligned} \qquad \ldots . (5)$$

In this last relation, each quantity viz. $v_0 - v_E$, and w, as we have seen, is of order m', and we are proposing to ignore all smaller terms involving $O(em')$. So in calculating w, any terms involving the eccentricity can be ignored, which means in effect that for the present purpose both Uranus and Neptune can be regarded as moving in circular orbits. But clearly, *when the orbits are circles the quantity w has equal and opposite values at equal times before and after conjunction*, that is for equal and opposite values of the time from t_0, that is at say $t_0 - \tau$. (The actual form of the expression for w is given later (p. 39) but is not explicitly required for the present purpose.)

So, using τ to denote time measured from conjunction,

$$w(t_0 - \tau) = -w(t_0 + \tau).$$

Or, if we write

$$W(\tau) = w(t) = w(t_0 + \tau),$$

then W has the important property that it is an *odd function* of τ, that is

$$W(-\tau) = -W(\tau). \qquad \ldots . (6)$$

Returning now to equation (5), the difference $v_0 - v_E$ represents the difference in longitude between two elliptic orbits whose elements differ by such amounts (of order m') as to make expression (3) applicable. Hence (5) becomes

$$\delta(t) = m'\{a + bt_0 + b\tau + c\cos(nt_0 + n\tau) + d\sin(nt_0 + n\tau)\} + W(\tau),$$

and in terms of new constants this can be written in the form

$$\delta(t_0 + \tau) = A + B(1 - \cos n\tau) + \{C\tau + D\sin n\tau + W(\tau)\}, \qquad \ldots . (7)$$

and it will be observed that the final expression in curled brackets is an *odd* function of τ. It is this feature that enables the time of conjunction to be found with considerable closeness. For this purpose, it follows at once from (7) that

$$\delta(t_0-\tau) = A+B(1-\cos n\tau)-\{C\tau+D\sin n\tau+W(\tau)\},$$

while

$$2\delta(t_0) = 2A,$$

and these combine to give the simple relation

$$\rho(\tau) \equiv \frac{\delta(t_0+\tau)-2\delta(t_0)+\delta(t_0-\tau)}{1-\cos n\tau} = B = \text{constant.} \qquad \ldots . (8)$$

If we denote the numerator here by $\Delta^2(\tau)$, then it represents a second difference of the observed discrepancies $\delta(t)$, and it is this form (8) that provides us with a criterion for determining conjunction. The rule it supplies is in words simply this:

> Select an instant t_0, calculate $\Delta^2(\tau) = \delta(t_0+\tau)-2\delta(t_0)+\delta(t_0-\tau)$ by means of Table 3 for a number of values of τ, divide each by the corresponding factor $(1-\cos n\tau)$, and the resulting quantity will be (approximately) constant if t_0 has been selected at conjunction. A series of values of t_0 will enable the time of conjunction to be found.

The constancy of $\rho(\tau)$ will of course be only approximate because of the terms in em', and others, neglected in arriving at (8).

5. Numerical Application of the Rule for Finding Conjunction

For this we make use of Table 3, and take for t_0 in turn the instants 1813·0, 1816·0, 1819·0, 1822·0, 1825·0, and 1828·0. Values of τ at intervals of 3 years were used since the Table gives these readily. The second difference $\Delta^2(\tau)$, as expected, is unsteady for small values of τ, especially $\tau = 3$, and to some extent $\tau = 6$, because of the errors in $\delta(t)$, as explained earlier. Accordingly the values for $\tau = 3$ are not utilized at all.

Having calculated $\Delta^2(\tau)$ and $\rho(\tau)$, the mean ρ of the latter quantities (weighted as explained below), is next found, and then the mean square difference (m.s.d.) of the values, $\rho(\tau)-\bar{\rho}$, is calculated to provide a measure of the constancy of $\rho(\tau)$ for each selected time of conjunction t_0. The resulting quantity, when graphed against t_0, shows an extremely marked minimum (just later than 1822·0) which can be read off to give the time of conjunction. The various quantities are tabulated in Table 4.

To take some account of the unsteadiness of $\Delta^2(\tau)$ for the smaller values of τ, the values of $\rho(\tau)$ were first given weights increasing outwards, thus 1 for $\tau = 6$, 2 for $\tau = 9$, 3 for $\tau = 12$, and so on. These weightings are shown under (a) in Table 4, and they lead to a certain value of the mean square difference. However, with increasing τ, the neglected terms in e are likely to have greater effect, and a second set of weightings were given at first increasing with τ and then decreasing to take some account of this. These weights are shown under (b) in Table 4. The result of taking equal weighting throughout is shown at (c). The calculations make it evident that the resulting time of conjunction is scarcely affected at all by the weighting factors, and the three curves shown in Fig. 3 corresponding to the different weightings demonstrate this.

All these quantities and the values of $\rho(\tau)$ are shown in Table 4 for a series of six values of t_0 spaced at intervals of 3 years. The function $\rho(\tau)$ is clearly most nearly constant for the year 1822 of those tabulated, and the graph of the mean square differences for the three weightings shows the minimum to come at about 1822·3 in all three cases. We have therefore arrived at the important result:

Predicted time of heliocentric conjunction = 1822·3.

As a matter of interest we may note at this point that the actual value was about six months earlier at 1821·74.

In the further calculations that are made it is convenient to take for the time of conjunction the exact beginning of a year so as more simply to use the observational material, and of course for the reason what we are here concerned more with method

Table 4

Values of $\rho(\tau)$ *at six selected instants,* 1813 *to* 1828, *to test its constancy. Minimum occurs near* 1822

	1813			1816			1819			1822			1825			1828		
τ	$\rho(\tau)$	a	b	$\rho(\tau)$	a	b	$\rho(\tau)$	a	b	$\rho(\tau)$	a	b	$\rho(\tau)$	a	b	$\rho(\tau)$	a	b
6	−0·25	1	1	8·99	1	1	23·69	1	1	42·27	1	1	88·39	1	1	108·4	1	1
9	4·08	2	2	10·52	2	2	27·58	2	2	53·12	2	2	78·80	2	3	95·14	2	3
12	6·10	3	3	15·83	3	3	35·64	3	3	53·12	3	3	74·70	3	3	87·05	3	2
15	11·20	4	4	23·62	4	4	38·56	4	3	54·64	4	2	70·87	4	1	—		
18	18·32	5	4	28·50	5	3	42·14	5	2	55·21	5	1	—			—		
21	23·66	6	3	33·59	6	2	45·13	6	1	—			—			—		
24	29·36	7	2	38·10	7	1	—			—			—			—		
27	34·65	8	1	—			—			—			—			—		
			m.s.d.			m.s.d.			m.s.d.			m.s.d.			m.s.d.			m.s.d.
	(a) 21·87		11·4	27·96		10·0	39·12		6·8	53·50		3·5	75·36		6·1	93·31		9·4
	(b) 15·43		10·2	22·67		9·4	35·91		6·9	52·48		4·1	77·47		5·7	96·63		10·0
	(c) 15·89		11·8	22·74		10·5	35·46		7·6	51·67		4·8	78·19		6·5	96·88		8·8

(The figures for 1813 can range 27 years on either side this date because the observational material, Table 1, extends to 1840, but in the next column for 1816 the range is reduced to 24 years, and so on. Thus there is one fewer entry in each succeeding column.)

(a) corresponds to weights increasing with τ,

(b) corresponds to weights increasing and then decreasing,

(c) corresponds to equal weights for each τ.

than with making extreme accuracy of prediction a century after it could be of direct practical value. It would be tempting, even for the demonstration of method, to choose the nearest year, viz. 1822·0, but this rounding off brings us closer to the actual time of conjunction. Accordingly, to avoid the possibility of improving things in this way, we will in fact adopt 1823·0 as the instant of conjunction, and proceed on this basis for the actual further calculations. (It would in fact be possible, as will be seen later, to introduce a small correction to the time of conjunction and solve for it, if extreme precision were our aim.) The initial error introduced by this later adopted time, as compared with 1822·0 say, will be the *difference* in the angular motions of Uranus and Neptune in 1 year, which in fact amounts to about 2°, and will start the unknown planet this far ahead. The actual predicted value for 1846·73 turns out to be about 1° *ahead* of the true position, the method (to be explained

below) evidently tending to correct somewhat, so that if 1822·0 were adopted and similarly investigated it would probably lead to a closer prediction, and in any event to a place not more than 1° *behind* the actual place.

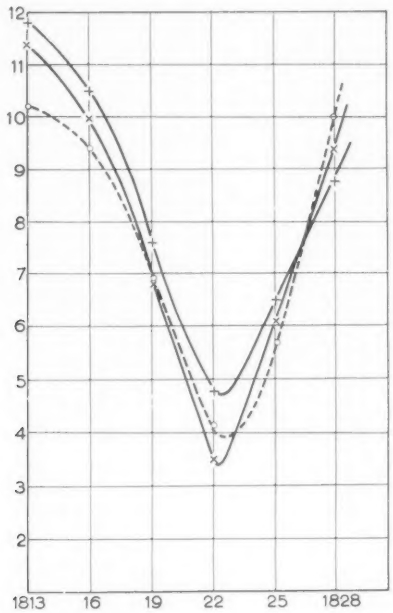


FIG. 3. Graphs showing degree of constancy of $\rho(\tau)$. The ordinate represents the root mean value of $(\rho-\rho)^2$ plotted at $t_0=1813$, 1816, 1819, 1822, 1825, and 1828. The different curves correspond to different weightings in Table 4. For each the minimum occurs at about 1822·3.

\+ equal weightings of $\rho(\tau)$
× weightings increasing with τ
○ weights increasing and then decreasing.

6. PRELIMINARY ASSESSMENT OF THE IMPORTANCE OF A KNOWLEDGE OF THE TIME OF CONJUNCTION

The time of conjunction having been fixed at 1823·0, and on the basis of Neptune moving in an accurately circular orbit, it is a simple matter to calculate its longitude for any assumed size of orbit on the date of discovery 23·73 years later. The angular rate of such a planet is proportional to $a'^{-\frac{3}{2}}$, where a' is its assumed mean distance, and the curve of Fig. 4 shows the resulting position for values of a'/a varying from about 1·4 to 2·1.

The curve of Fig. 4 shows at once that a range of only $\pm 10°$ in longitude on either side of the discovery position is covered by the relatively very great range of a'/a from about 1·43 to 1·89. Even on the simple assumption of BODE's law, viz. $a'/a \doteq 2$, the predicted longitude is only about 13° behind the actual position. Thus the determination of the time of conjunction, for the particular circumstances of Neptune at the time, would in itself have sufficed to give a fair indication of the region of the sky to be searched. At AIRY's suggestion, CHALLIS actually decided to take a zodiacal

belt 10° wide and extending 30° in longitude. If this had been centred on the place here indicated by BODE's law (with the present approximate time of conjunction), the planet would have been well inside the area, and eventual discovery quite possible. The use of more observations than here made use of might well give a better determination of conjunction and possibly reduce the error of prediction (by BODE's law) to about 10°.

In Fig. 4, the limits of distance (in a circular orbit) are also indicated that correspond to the errors of the predicted position made by the methods actually used,

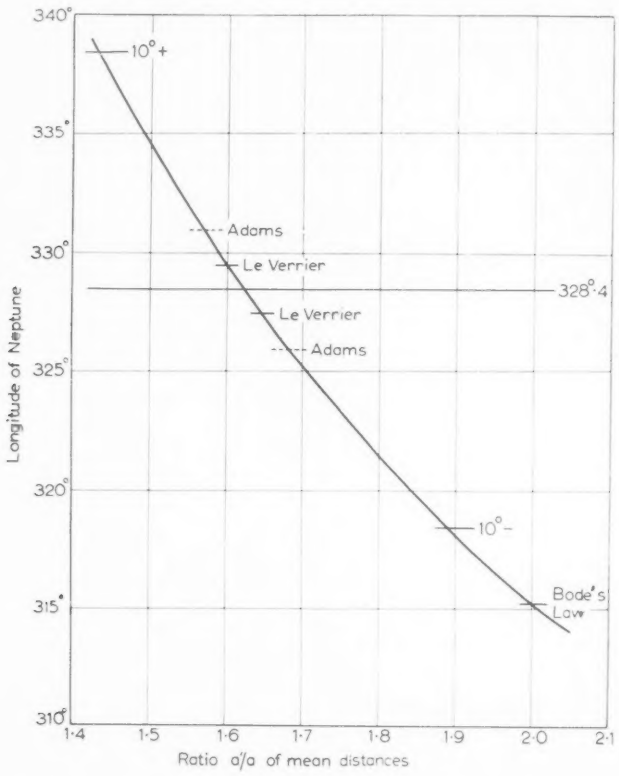


FIG. 4. Curve showing the predicted positions of Neptune at the date of discovery (1846·73), based on conjunction at 1823·0, for values of the distance ratio a'/a ranging from 1·4 to 2·1. The long horizontal line shows the actual longitude of Neptune at discovery (328°·4), and the short lines the permissible limits of a'/a for errors not exceeding those made by ADAMS, by LE VERRIER, and for an error within ±10° of longitude.
[LE VERRIER (1° error), range 1·60 to 1·645; ADAMS ($2\frac{1}{2}$° error), range 1·57 to 1·68; ±10° of longitude, range 1·43 to 1·89].

which amounted to about $2\frac{1}{2}$° for ADAMS's calculations, and to about 1° for those of LE VERRIER. These limits are marked on the curve by short horizontal lines. A prediction error not exceeding that of ADAMS is seen to be possible if the ratio of distance a'/a could be deduced from the observations, by further calculation, within the range 1·57 to 1·68, while the closer prediction of LE VERRIER would be covered by a range of about 1·60 to 1·645.

The question accordingly arises, in view of the serious failure on this one important occasion of BODE's law, which had it held might have rendered discovery so easy, whether by simple methods not involving arduous arithmetic a close estimate of the mean distance of Neptune can be made from the observational material (Tables 2 and 3). It will appear that this is in fact entirely possible, and a result is obtained that agrees within about 1° of the actual longitude. We proceed to explain the means of achieving this, but it will help to explain the method if beforehand we discuss what was involved in the method of LE VERRIER, and in principle also that of ADAMS.

7. WHY ADAMS'S AND LE VERRIER'S METHODS WERE SO EXTENSIVE IN APPLICATION

With only a best-fitted ellipse to represent the position of Uranus, the accurate analytical expression for its position (assuming coplanar orbits in the ecliptic for both Uranus and the unknown planet) will in fact involve *nine* unknowns. First of all the accurate but unknown elements of the elliptic orbit of Uranus after the removal of *all* perturbations will mean four unknowns, which in fact can be considered as small corrections Δn, $\Delta\epsilon$, Δe, and $\Delta\varpi$ to the known approximate elements of m, and so entering the equations only linearly. Over and above these will occur the elliptic elements of the unknown planet, namely a' (which settles n'), ϵ', e', and ϖ', which together with its mass m' make nine in all. Once ADAMS and LE VERRIER decided on a non-circular orbit for m', the introduction of the elements e' and ϖ' could not be avoided, and conjunction not being known, ϵ' also has to be introduced. It is these that cause the great difficulty.

The natural object of any process for determining such quantities as these is to arrive in some way at linearised equations that can be treated by least-squares. As far as the corrections to the orbit of Uranus are concerned, these will enter linearly. But as to the others, the form that the principal perturbations of m due to m' take is given by:[4]

$$P(t) = m'\sum_{1}^{\infty}F_i \sin i\{(n'-n)t+\epsilon'-\epsilon\}+m'e\sum_{-\infty}^{\infty}G_i \sin[i\{(n'-n)t+\epsilon'-\epsilon\}+nt+\epsilon-\varpi$$

$$+m'e'\sum_{-\infty}^{\infty}H_i \sin[i\{(n'-n)t+\epsilon'-\epsilon\}+nt+\epsilon-\varpi'], \qquad \ldots . (9)$$

to the first order in m' and in the eccentricities. In this expression, the coefficients F_i, G_i, H_i are functions of the ratio of the mean distances, and have elaborate forms. For instance (TISSERAND, *loc. cit.* p. 366)

$$F_i = \frac{i(z_i^2+3)}{z_i^2(1-z_i^2)}aA^{(i)}+\frac{2z_i}{z_i^2(1-z_i^2)}a^2\frac{\partial A^{(i)}}{\partial a}, \qquad \ldots . (10)$$

wherein $z_i = i(1-n'/n)$, (i taking successive integer values), and

$$A^{(i)} = \frac{2}{a'}\cdot\frac{1.3.5\ldots(2i-1)}{2.4.6\ldots.2i}\left(\frac{a}{a'}\right)^i\left[1+\frac{1}{2}\cdot\frac{2i+1}{2i+2}\left(\frac{a}{a'}\right)^2\right.$$

$$\left.+\frac{1.3}{2.4}\cdot\frac{(2i+1)(2i+3)}{(2i+2)(2i+4)}\cdot\left(\frac{a}{a'}\right)^4+\ldots\right], \qquad \ldots . (11)$$

[4] *F. Tisserand, Traité de Mécanique Céleste*, vol. I, p. 365, Paris 1889.

with similar expressions in infinite series for G_i and H_i. Accordingly, unless a definite numerical value is assumed for a/a', there is no possibility of arriving at linear equations. This plainly is why both ADAMS and LE VERRIER were tempted to adopt BODE's law. But even with a/a' settled, rightly or wrongly, there still remain *eight* unknowns, and quite apart from the large task of solving the corresponding equations of conditions, not all these remaining eight quantities occur linearly. The principal difficulty arises from the unknown ϵ'. If as is usual we write

$$h' = e' \cos \varpi', \quad k' = e' \sin \varpi',$$

then (9) may be put in the form

$$P(t) = Fm' + Hm'h' + Km'k' \qquad \ldots . (12)$$

wherein now F, H, and K depend *trigonometrically* and not linearly on the *unknown* ϵ', sines and cosines of ϵ', $2\epsilon'$, $3\epsilon'$, and so on, entering according to how many terms of (9) are retained.

To meet this difficulty, LE VERRIER first reduced the number of observations to be used to 18, taking means to arrive at best values, spaced out over the period from 1690·98 to 1845·70. He then in turn took 40 values of ϵ' at 9° intervals, so that the whole possible range from 0° to 360° was covered. For each assumed value of ϵ', the expression for $P(t)$ immediately becomes linear if the unknowns are regarded as m', $m'h'$, and $m'k'$, and it is then possible to solve the equations of condition by least-squares. Having carried out *forty* such solutions, LE VERRIER then selected that one giving the best fit to the whole series of observations, and in this way arrived at the value $\epsilon' = 252°$ as the best value. Once such an approximately satisfactory value of ϵ' became available, it could of course be improved upon by the further use of linearised equations. This was LE VERRIER's method, but clearly much arithmetical labour is involved in such an approach to the problem. It is proposed then to show next how these awkward features can in fact be obviated once the instant of conjunction is known, and supposing always that the objective is simply prediction of the position by means of circular orbits only.

8. METHOD FOR THE DETERMINATION OF THE MEAN DISTANCE OF THE UNKNOWN-PLANET

The assumption of a circular orbit, together with a knowledge of the time of conjunction, immediately reduces the number of unknowns associated with Neptune to two, namely its mass m' and its constant heliocentric distance a', and removes altogether the extremely awkward feature of the appearance of ϵ' trigonometrically, just explained, since once conjunction is known the time can be measured from it. Also, in (9), since the eccentricity of the orbit of Uranus $e = 0{\cdot}047$, the terms in $m'e$ are small compared with those from the part in m' alone. Accordingly, though this is by no means essential to the method, we propose to calculate the perturbations as though Uranus itself was moving in a circular orbit. This effects a certain reduction in the computations and is sufficient for the demonstration of the method. The terms in e can in fact readily be included for higher accuracy.

The whole expression for the correction to the longitude of m will then consist of

two parts: first there will be that depending merely on the corrections required to the elements of its best-fitting ellipse, which part takes the form

$$\Delta v = \delta\epsilon + t\delta n + 2\sin M.\,\delta e - 2\cos M.\,e\delta\varpi, \qquad \ldots\ (13)$$

where M is the mean anomaly $nt+\epsilon-\varpi$. This expression is unaltered in form whatever origin is chosen for t, so that we have always

$$\Delta v = \delta\epsilon + t\delta n + \delta\alpha\sin nt + \delta\beta\cos nt, \qquad \ldots\ (14)$$

where $\delta\alpha$ and $\delta\beta$ are linear combinations of the original corrections to the elements. Accordingly, if we agree to measure t from conjunction, the whole expression for the discrepancies in heliocentric longitude will be given by

$$\delta v = \delta\epsilon + t\delta n + \delta\alpha\sin nt + \delta\beta\cos nt + P(t) \qquad \ldots\ (15)$$

where now the perturbations reduce simply to

$$P(t) = m'\sum_{i=1}^{\infty} F_i \sin i(n-n')t. \qquad \ldots\ (16)$$

This last expression is in fact the analytical form of the odd function w referred to (p. 32) earlier in obtaining the time of conjunction.

If now a direct comparison were made between (15) and the observations, with an assumed value for a' (and hence of n' related to it by $n'^2a'^3 = 1$), to give a set of equations of condition, the number of unknowns involved would be *five*, namely $\delta\epsilon$, δn, $\delta\alpha$, $\delta\beta$, and m', which is a considerable reduction on the number involved in the methods of ADAMS and LE VERRIER. But it would still be necessary to try a series of value of a' in order to get the best fit. In view of this, it is convenient first to introduce a simple device of regrouping the equations of condition that has the effect of eliminating $\delta\alpha$ and $\delta\beta$.

To see how this can be done, consider the trigonometrical identities

$$\begin{aligned}\sin(x+\beta)+\sin(x-\beta)-\sin x &= (2\cos\beta-1)\sin x,\\ \cos(x+\beta)+\cos(x-\beta)-\cos x &= (2\cos\beta-1)\cos x.\end{aligned} \qquad \ldots\ (17)$$

If in these β is taken as 60°, the right-hand sides vanish. Now in the orbit of Uranus, the period $2\pi/n$ is 84·013 years, and accordingly 60° is described in almost exactly 14 years. Hence, if instead of a particular observation $f(t)$ say, we adopt $f(t+14)+f(t-14)-f(t)$, the time t of course being measured now in years, then the terms in $\delta\alpha$ and $\delta\beta$ in (15) will be eliminated. This step is of course no more than a regrouping of the equations of condition (reducing their number), but as will appear later it also has a certain other important effect where orbits near the resonance $n'/n = 2$ are concerned.

Under this regrouping, the terms $\delta\epsilon+t\delta n$ come out the same. But the terms in $P(t)$ have to be adjusted. If we put $D = (n-n')t$, then we may write

$$P(t) = m'(F_1\sin D + F_2\sin 2D + F_3\sin 3D + F_4\sin 4D + \ldots) \qquad \ldots\ (18)$$

wherein it turns out that for any ratio a/a' of interest here the coefficients $F_1, F_2, F_3, \ldots$ diminish fairly quickly with increasing suffix (with the exception of that of $\sin 2D$ near the resonance $n = 2n'$ where it can become very large). Accordingly, where the new equations of condition are concerned, the appropriate value of β involved in (17),

since now the corresponding $x = (n-n')t$ for the term in sin D, and twice this for sin $2D$, and so on, will be given by

$$D+\beta = (n-n')(t+14) \qquad \text{where } n\,.\,14 = 60^\circ,$$
$$= D+(1-n'/n)60^\circ.$$

So, writing $\nu = n'/n$ for the ratio of mean motions, $\beta = (1-\nu)60^\circ$. Accordingly to transform $P(t)$, successive coefficients in (18) must be multiplied by the following factors:

$$2\cos\{(1-\nu)\ 60^\circ\}-1 \text{ for } \sin D,$$
$$2\cos\{(1-\nu)120^\circ\}-1 \text{ for } \sin 2D,$$
$$2\cos\{(1-\nu)180^\circ\}-1 \text{ for } \sin 3D,$$

and so on, where $\nu = n'/n$.

With these conversion factors applied to the terms of $P(t)$, we may denote the resulting expression by $Q(t)$, so that in fact

$$Q(t) = P(t+14)+P(t-14)-P(t). \qquad \ldots\ (19)$$

and we can write the resulting series of terms formally as

$$Q(t) = m'(f_1 \sin D+f_2 \sin 2D+f_3 \sin 3D+\ \ldots). \qquad \ldots\ (20)$$

Accordingly, the revised equations of condition replacing (15) take the form

$$\delta\epsilon+t\delta n+Q(t) = \delta v(t+14)+\delta v(t-14)-\delta v(t), \qquad \ldots\ (21)$$

and now for given a', and hence a/a', only *three* unknowns are involved, namely $\delta\epsilon$, δn, and the mass m' which occurs in $Q(t)$.

The procedure we shall adopt then will be to set up these equations of condition for a series of values of a/a' to find what value of the ratio gives the best fit of the equations when solved by least-squares. The first step is to transform the observations themselves, and the following Table gives the adjusted values, which follow immediately from Table 3.

Table 5

Values of $\delta v(t+14)+\delta v(t-14)-\delta v(t)$

	1794	5″61	*	1805	18″30		1816	+ 2″00
	95	6″75		06	18″65	*	17	− 3″99
*	96	8″30		07	18″95		18	− 9″63
	97	9″40	*	08	18″82		19	−15″25
	98	10″55		09	18″55	*	20	−20″45
*	99	12″00		10	18″30		21	−27″13
	1800	13″05	*	11	17″21		22	−33″58
	01	14″10		12	16″25	*	23	−40″21 (conjunction)
*	02	15″25		13	14″10		24	−47″51
	03	16″41	*	14	10″57		25	−54″06
	04	17″48		15	7″80	*	26	−60″89

Comparing this with Table 3, it will be noticed that this regrouping involves the loss, as it were, of 14 years at each end of the series of observations, simply because of the $t+14$ and $t-14$ terms.

In carrying out the task of finding the value of a' that gives the best solution of

equations (21) for the adopted value of $t_0 = 1823{\cdot}0$, it is a further economy of arithmetic if we set up the equations only at 3-year intervals, so that in fact we shall make use only of the eleven starred entries in Table 5. This also is in keeping with the nature of the original observational material of Table 2 which itself refers only to 3-year intervals.

9. Calculation of $Q(t)$

In order to arrive at $Q(t)$, the numerical value of the coefficients F_i of $P(t)$ in (13) are first needed, and these have been found for a series of values of the ratio a/a' appropriate to the range that will be of interest.

Table 6

Coefficients in the series for $P(t)$

$\alpha = a/a'$	sin D	sin $2D$	sin $3D$	sin $4D$
0·50	− 5·738	1·429	0·141	0·031
·55	− 7·432	3·380	0·299	0·068
·60	− 9·911	15·712	0·672	0·154
·65	−13·718	−38·422	1·692	0·366
·70	−19·944	−18·655	5·461	0·967

Each of these has now to be adjusted by multiplying by the appropriate value of the factor $(2\cos\beta - 1)$ to get the corresponding coefficients f_i in $Q(t)$ given by (20). The resulting values are as follows:

Table 7

Coefficients in the series for $Q(t)$

$\alpha = a/a'$	sin D	sin $2D$	sin $3D$	sin $4D$
0·50	− 3·208	−0·814	−0·266	−0·087
·55	− 4·664	−1·290	−0·469	−0·176
·60	− 6·877	−2·052	−0·821	−0·345
·65	−10·382	−3·299	−1·436	−0·666
·70	−16·165	−5·471	−2·558	−1·284

In the above values for $P(t)$, the coefficient of sin $2D$ becomes large, as Table 6 to some extent shows, and changes sign as $\alpha = a/a'$ passes through the resonance value 0·6294 corresponding to $n = 2n'$. But it is to be noticed that the conversion factor $2\cos\beta - 1$ vanishes at this resonance, and evaluation of the limit of $(2\cos\beta - 1)F_2$ in fact shows that the corresponding coefficient in $Q(t)$ runs smoothly through the resonance without singularity (though the form (18) does not of course apply if $n = 2n'$ accurately). Accordingly, the adjusted function $Q(t)$ is of such a form, as Table 7 shows, that for other values of α than those for which the coefficients have actually been calculated it can be written down with sufficient accuracy by simple interpolation. The range we shall be concerned with is roughly from about $\alpha = 0{\cdot}5$ to $\alpha = 0{\cdot}63$, and it will be seen from Table 7 that there is a clear trend in the coefficients making fairly accurate interpolation easily possible. It turns out in fact that extreme

accuracy in the coefficients is not of great account. If we think of the expression for $Q(t)$ as a curve, then its general form gradually changes as α diminishes, and our aim is to find what α gives the closest fit to the (adjusted) observations.

It is in fact also possible to include the terms of (9) with factor $m'e$ (since the elements of the orbit of Uranus may be regarded as known for this purpose) either by directly evaluating them and including the largest terms, or by representing the principal terms as themselves expanded in a sine series in terms of D for the limited arc of the orbit of Uranus, to which the observations relate. By this second method, they then appear as slight additions to the coefficients in the sine series for $Q(t)$. In either case, the general form of the curve defined by $Q(t)$ is unaffected because even the principal terms in $m'e$ turn out to be small and comparable only with the coefficients of the higher terms in $Q(t)$.

10. Solution of the Equations of Condition

The equations of condition have been set up and solved for a series of values of a'/a to find that value at which the best fit occurs. The values actually selected were

$$a'/a = 2{\cdot}0,\ 1{\cdot}8225,\ 1{\cdot}69,\ 1{\cdot}6129,\ 1{\cdot}595169,\ 1{\cdot}5625,\text{ and } 1{\cdot}53846,$$

and it has been found that the least squares solution shows a very pronounced best fit at just about $a'/a = 1{\cdot}6$. (See Fig. 5 below.)

We give here details of the equations of condition corresponding to (21) for the case $a'/a = 1{\cdot}595169$ which turns out to be nearest the best fit. The form of the function $Q(t)$ is found to be (in radians)

$$Q(t) = -8{\cdot}5\frac{m'}{\odot}(\sin D + 0{\cdot}309\sin 2D + 0{\cdot}129\sin 3D + 0{\cdot}058\sin 4D + \underline{0{\cdot}029}\sin 5D + \underline{0{\cdot}016}\sin 6D). \quad \ldots\ (22)$$

In this expression, the last two coefficients (underlined) have been included as estimates since their approximate values so readily follow from the trend of the earlier coefficients, and they have similarly been included in the solutions for other values of a'/a not given here in detail.

To change to seconds of arc, for comparison with the actual observations, the factor 206,265″ must be introduced, and we write

$$\mu = 8{\cdot}5\frac{m'}{\odot} \times 206265 \quad \ldots\ (23)$$

the parameter μ replacing m' with a view to arriving at reasonably sized coefficients in the equations of condition. Once μ is found, (23) will then give the mass m' of the unknown planet. The equations of condition may be set out schematically in a form shown in Table 8:

The figures in the 2nd, 3rd, and 4th columns are the coefficients of $\delta\epsilon$, δn, and μ, with the unit of time 1 year and its origin at conjunction 1823·0. When these are solved by least squares, the corresponding best values are found to be

$$\left.\begin{aligned}\delta\epsilon &= -40{\cdot}61\\ \delta n &= -\ 0{\cdot}384\\ \mu &= \ \ 68{\cdot}5\end{aligned}\right\} \quad \ldots\ (24)$$

Table 8

Equations of Condition for $a'/a = 1{\cdot}595169$

Year	$\delta\epsilon$	δn	μ	Obs.	Calc.	O—C
1826	1	3	−·276	−60·89	−60·67	−·22
1823	1	0	0	−40·21	−40·61	·40
1820	1	− 3	·276	−20·45	−20·55	·10
1817	1	− 6	·514	− 3·99	−3·12	−·87
1814	1	− 9	·684	10·57	9·73	·84
1811	1	−12	·776	17·21	17·14	·07
1808	1	−15	·793	18·82	19·50	−·68
1805	1	−18	·756	18·30	18·13	·17
1802	1	−21	·691	15·25	14·78	·45
1799	1	−24	·638	12·00	12·32	−·32
1796	1	−27	·562	8·30	8·28	·02

By means of these, the calculated values as shown in the 6th column can then be found, and thence the difference O − C. It is then readily found from these that for the mean square difference $\Sigma(\mathrm{O}-\mathrm{C})^2/8 = 0{\cdot}31$.

For different values of a'/a, correspond different coefficients of μ, since n' is changed and also the coefficients of the sine terms in $Q(t)$. For the values of a'/a investigated, the following results were obtained:

Table 9

a'/a	2·0	1·8225	1·69	1·6129	1·595169	1·5625	1·53846
$\frac{1}{8}\Sigma(\mathrm{O}-\mathrm{C})^2$	2·65	1·71	1·49	0·37	0·31	5·97	9·46
m.s.d.	1·63	1·31	1·22	0·61	0·56	2·44	3·08

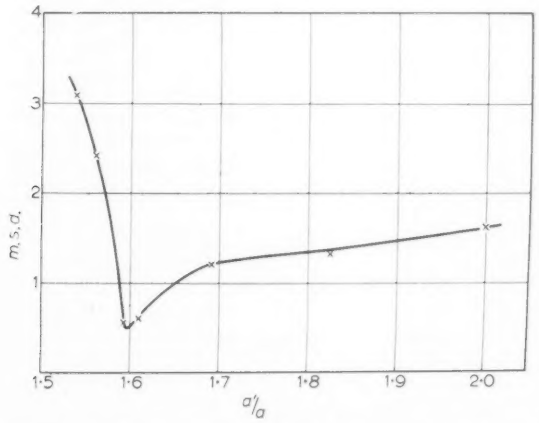


FIG. 5. Curve of mean square deviation (m.s.d. of Table 9) computed for a number of values of a'/a. The best fit is seen to occur at about $a'/a = 1{\cdot}6$.

The curve of the m.s.d. plotted against a'/a is shown in Fig. 5. As a'/a decreases, these values show the strong minimum in $\frac{1}{8}\Sigma(\mathrm{O}-\mathrm{C})^2$, already referred to, occurring at just about $a'/a = 1{\cdot}6$. Since the residuals must themselves be subject to error, there is no point in attempting to determine the position of minimum more closely than the round figure $a'/a = 1{\cdot}6$ indicated by the above values. Returning with this value to Fig. 4, it is seen at once that the result places the unknown planet at just about

the same accuracy as LE VERRIER's limit, and in fact about 1° ahead of its actual position at discovery.

Adopting $a'/a = 1{\cdot}6$ as our final result, it will be sufficient for the calculation of m' to use the value of μ obtained for $a'/a = 1{\cdot}595169$ as given in (24). By means of (23) we then have

$$8{\cdot}5 \times \frac{m'}{\odot} \times 206265 = 68{\cdot}5$$

or $m'/\odot = 1/25{,}500$. For the mean distance, since $a = 19{\cdot}191$ a.u., we have $a' = 30{\cdot}71$ a.u. These values, with those obtained by ADAMS and LE VERRIER and the actual ones for Neptune, give the following comparison (Table 10 and Fig. 6).

Table 10

	ADAMS	LE VERRIER	Present method	Neptune
Mean distance	38·4	36·2	30·71	30·07
Eccentricity	0·161	0·108	0·000	0·0086
Longitude of perihelion	316°	284°	—	44°
Sun/mass	6040	9350	25,500	19,300
Heliocentric Longitude at discovery (equinox 1950)	330°·9	327°·4	329°·4	328°·4

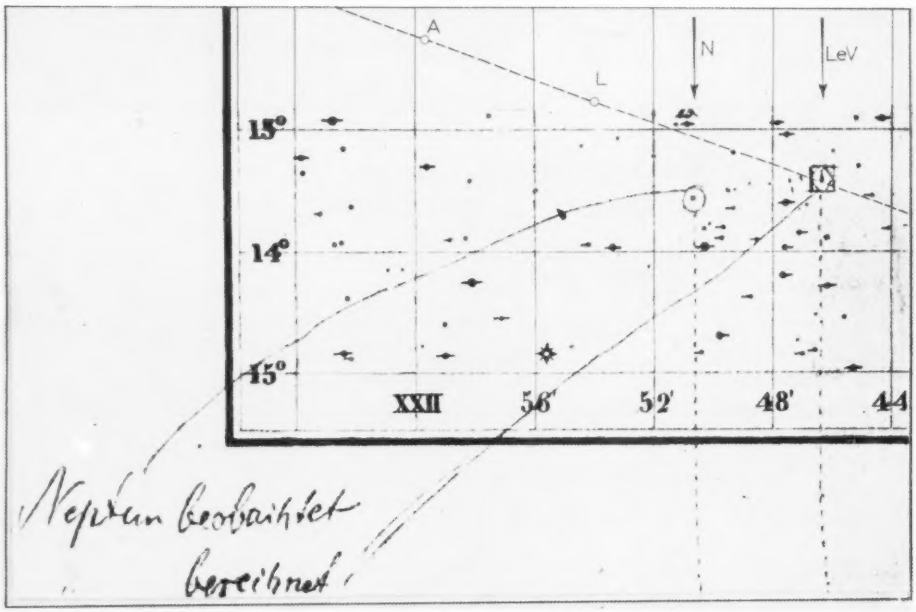


FIG. 6. Section of the original chart ("Akademische Sternkarte" by BREMIKER, equinox 1800·0) which enabled GALLE in 1846 to find Neptune (N) using LE VERRIER's prediction (LeV). The brightest star in this field (at R.A. 21^h55^m6 and Dec. $-14°50'$) is ι Aquarii, mag. 4^m3. The entries referring to Neptune's observed and calculated positions (bottom left) are supposed to have been made by ENCKE. In this reproduction, the ecliptic has been added, and the coordinates have been extended to Dec. $-12°$, in order that ADAMS's prediction (A) and the result of the present paper (L) may be shown.[5]

[5] I am indebted to Dr. A. BEER for this diagram. R.A.L.

11. The Methods of Adams and E. W. Brown

In concluding this paper we may mention first what appears to be a surprisingly obvious error that occurs at the very outset of Adams's method. (It is because of the presence of this in Adams's work that I have preferred to discuss the method of Le Verrier by way of comparison.) The error to which I refer is in my view embodied in the procedure described by Adams (*loc. cit.*, Vol. I, p. 11) as follows:

> "It is easily seen that the series expressing the corrections of *mean* longitude in terms of the corrections applied to the elements of the orbit, is more convergent than that which gives the correction of the *true* longitude, and the same thing is true for the perturbations of the mean longitude, as compared with those of the true. The corrections found above were accordingly converted into corrections of mean longitude by multiplying each of them by the factor r^2/ab, r being the radius vector, and a and b the semi-axes of the orbit (of Uranus)."

Now whereas such a relation as that claimed subsists between the true longitude and the mean longitude in an *undisturbed* elliptic path in virtue of the equation

$$r^2 \frac{dv}{dt} = h = nab,$$

which may be written

$$(r^2/ab)dv = ndt = dl,$$

where l denotes the mean longitude, there seems no validity in supposing that in *perturbed* motion the unexplained variation in true longitude Δv and that in mean longitude Δl are similarly connected. No such relation seems known in celestial mechanics, and indeed if it existed, the conversion made by Adams could only mean multiplying both sides of an equation of condition by the same factor, viz. r^2/ab (whose value presumably would refer to the instantaneous orbit at time t). But in fact we have in any case

$$v = \int hr^{-2}dt, \qquad l = nt + \epsilon,$$

so for perturbations (Δ) in which all the elements are changing

$$\Delta v = \int (r^{-2}\Delta h - 2hr^{-3}\Delta r)dt, \qquad \Delta l = t\Delta n + \Delta\epsilon,$$

and hence

$$\Delta l/\Delta v = (t\Delta n + \Delta\epsilon)/\int (r^{-2}\Delta h - 2hr^{-3}\Delta r)dt.$$

In the denominator of this expression, Δh depends on Δa and Δe, while Δr depends on Δa, Δe, $\Delta\epsilon$, and $\Delta\varpi$, so there seems no possibility of this reducing to any form independent of all these increments.

In the hope of discovering a justification of this procedure, I have referred to a number of dynamical astronomers but without success. Indeed, I have recently found that no less a person than E. W. Brown was puzzled by this step, but in the absence of adequately detailed explanation in Adams's own account of the matter, E. W. Brown[6] contents himself with saying that "the meaning of this process is

[6] E. W. Brown, "On a Criterion for the Prediction of an Unknown Planet", M.N. **92**, 80, 1931.

not clear", and expressing the possibility that "the change may have been taken care of in some other way". In fact when applied to the residuals (Table 2) to be explained, for the most part it means multiplication by a factor very near unity, with the result that the general form of the curve of residuals (against time) of Fig. 2 is not much changed, and presumably the mean longitude residuals would show a similar form when properly found from those of true longitude, but the last three residuals are changed from $-20{\cdot}80$, $-42{\cdot}66$, $-66{\cdot}64$ to $-22{\cdot}63$, $-46{\cdot}70$, $-73{\cdot}09$, and it may be that the large changes so introduced are erroneous and account in some way for ADAMS's final result of the predicted place being poorer in accuracy than that of LE VERRIER. But the fact that he nevertheless succeeded to the extent he did illustrates the same point made in the second part of this paper, namely that the general form of the curve of residuals, rather than the absolutely precise values, is almost sufficient to determine with fair accuracy the position of the planet within a reasonably short time after conjunction.

The same point is emphasized by E. W. BROWN himself,[6] where he outlines a method similar to that used here but without the initial step of determining otherwise the time of conjunction. The method proposed by him rests entirely on consideration of the series $-m'\Sigma F_i \sin iD$ for the principal perturbations in longitude. Bearing in mind that it is the general form of the curve this represents that is of importance, E. W. BROWN proceeds on the basis that the first three terms will suffice for a rough prediction. However, I am doubtful of the numerical accuracy of the resulting work since I am unable to recover the numerical values he gives for many of the various coefficients, and these doubts are increased by his conclusion that the method gives conjunction as "close to 1840". This, for Uranus, is nearly 90° from the true position, and it would lead to a prediction for the longitude of the unknown planet at the time of discovery (even assuming a circular orbit at exactly the right distance) some 27° ahead of the actual place. Nevertheless, it would appear from the present investigation that the general basis of E. W. BROWN's suggested method is sound, though starting as it does without a definite knowledge of the time of conjunction, otherwise determined, its application with sufficient accuracy would involve considerably greater numerical labour than the method described herein.

The Determination of Einstein's Light-Deflection in the Gravitational Field of the Sun

H. von Klüber

The Observatories, Cambridge, England

Summary

The problem of the observational determination of the light-deflection in the gravitational field of the Sun, as predicted by Einstein's General Theory of Relativity, is outlined. All available results obtained at eclipse expeditions until now, as far as these have been successful, and their critical discussions are briefly summarized. Each set of observations is represented diagrammatically by the particular star field, and the published measures are shown for each star separately. The relevant details of each attempt are tabulated. An extensive bibliography covers most of the essential work on the problem.

1. Historical Introduction

According to classical theory, in empty space light should be propagated in a straight line. At the beginning of the 19th century (when there was still some dispute as to whether light should be considered as corpuscular or purely as a wave-motion), Soldner (1801) investigated the behaviour of a light-ray in a gravitational field of the classical Newtonian type, assuming the corpuscular theory. Unfortunately, his formula contains the erroneous factor 2. Correcting for this, and using modern constants, it can be shown that light coming from a star, and just grazing the limb of the Sun before reaching an observer on the Earth, should be deviated by an angle of $0''.87$.

In 1908, and again in 1911, before he had developed his General Theory of Relativity, and probably without knowing anything of Soldner's earlier work, Einstein investigated a possible deflection of light in a gravitational field (Einstein, 1908, 1911). Starting from the conception of the equivalence of a uniform gravitational field and an accelerated system of reference, he arrived at the conclusion that radiation energy (light, etc.) must have inertia or mass, and that this mass must be subject to gravitational forces. From the principle of equivalence he derived directly a formula for the deviation of light (as it should appear to a terrestrial observer looking at a star, the light of which had passed through a Newtonian gravitational field near the Sun). He found the expression

$$\alpha = \frac{2kM}{c^2 r}, \qquad \ldots . (1)$$

where the angle of deflection is α (Fig. 1); k is the constant of gravitation, M the mass of the Sun, r the distance at which the light ray passes the centre of gravity, and c the speed of light in vacuum. For the limb of the Sun he found the value $\alpha = 0''.87$. This, as is to be expected, is the same value as that found by Soldner, although derived in a far more general way.

A first, somewhat tentative attempt to prove the existence of such an effect by an examination of older plates, taken for other purposes, gave no conclusive results

(FREUNDLICH, 1913). Special observations planned by the Cordoba Observatory at the eclipse of 1912, probably the first expedition aiming at the determination of the light-deflection, failed because of bad weather (PERRINE, 1923). A number of further suggestions were made to check the expected light-deflection by special observations during total solar eclipses (FREUNDLICH, 1913; CURTIS, 1913; FREUNDLICH, 1914). Even a proposal of making day-time observations for this purpose was thoroughly discussed (LINDEMANN, 1916a, 1916b). Later on, also the possibility of making suitable observations of the major planets was considered (TRUMPLER, 1929a, 1929b). Another attempt for observing the light-deflection in 1914 was prevented because of the outbreak of war; (FREUNDLICH, 1930). Shortly afterwards EINSTEIN (1916) published his famous General Theory of Relativity, in which he used his new law of gravitation, differing from NEWTON's classical law by small terms; it nevertheless become appreciable for the light-deflection close to the Sun. His formula reads

$$\alpha = \frac{4kM}{c^2 r} \qquad \ldots . (2)$$

and with the modern values

$$k = 6{\cdot}67 \,.\, 10^{-8}\ \mathrm{cm^3 g^{-1} sec^{-2}}$$

$$M = \text{mass of the Sun} = 1{\cdot}991 \,.\, 10^{33}\mathrm{g}$$

$$c = 2{\cdot}998 \,.\, 10^{10}\ \mathrm{cm\ sec^{-1}}$$

$$r = \text{radius of the Sun} = 6{\cdot}956 \,.\, 10^{10}\ \mathrm{cm},$$

we obtain for α (at the Sun's limb) the usually quoted maximum value $L = 1''{\cdot}75$. Another calculation of this value was given by M. v. LAUE (1920). A confirmation of the formula (2) would obviously provide a crucial test for the whole concept of EINSTEIN's General Theory. As the formula cannot be verified by any experiment in a terrestrial laboratory, its confirmation by adequate astronomical observations is of fundamental importance.

Such work has so far been possible only during the brief and rare opportunities offered by total solar eclipses, because only then can we observe some stars sufficiently near to the Sun's limb.

At the 1918 eclipse a Lick Observatory expedition succeeded in obtaining photographs showing up to 50 stars around the Sun. But since the special technique required for these very difficult observations and their reduction was not yet fully developed, no conclusive results were obtained (CAMPBELL, 1919). Since then, a large number of further attempts were made by different observers. Owing to the great technical difficulties and the scarcity of solar eclipses, only very few of these have been even moderately successful, and more accurate observations, especially within one solar radius beyond the limb, are still urgently needed.

Actual results from observations of the light deflection in the gravitational field of the Sun have so far been published only for six eclipses. This is not surprising, for two reasons. As can be seen from what follows, these observations must be considered still as amongst the most difficult which can be attempted at an eclipse. Furthermore, from 1916 when EINSTEIN first published his formula (2) up to 1958 only 24 total solar eclipses have taken place, giving altogether a total observing time of not more than about 90 minutes. (The longest possible duration of a total

solar eclipse is about $7\frac{1}{2}$ minutes (LEWIS, 1931) and such an occasion occurs very seldom.) A considerable number of these 24 eclipses could not be used, because either the star field surrounding the Sun was unsuitable, or the duration of the eclipse was too short; bad weather at the moment of the eclipse, and uncertain political circumstances have prevented any observations at quite a number of other eclipses.

During earlier years there was much discussion and controversy as to whether the predicted Einstein effect may be influenced, or masked, or simulated by other physical factors in the near surroundings of the Sun; for instance, by diffraction in the

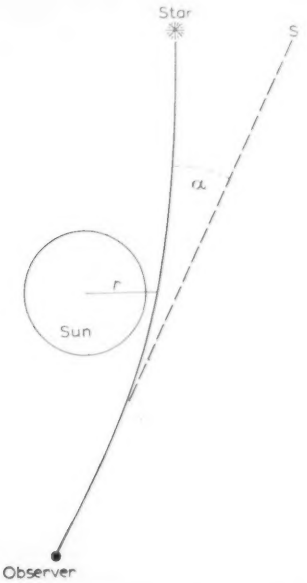


FIG. 1. The light-deflection as seen by a terrestrial observer. The General Theory of Relativity predicts a shift of a star image by the angle α, away from the Sun, according to equation (2). If the light ray grazes the Sun's limb, α reaches its maximum 1".74; this value is usually quoted as the "Einstein Effect", denoted in this paper by L.

Sun's atmosphere, or even by effects in the Earth's atmosphere (JONCKHEERE, 1918; LINDEMANN, 1918; EDDINGTON, 1918a, 1918b, 1918c; ANDERSON, 1919, 1920a, 1920b, 1922, 1924; DINES, RICHARDSON and ANDERSON, 1918; JEFFREYS, 1919; Royal Astronomical Society discussion, London, 1919a, 1919b; EDDINGTON and CROMMELIN, 1919; NEWALL, 1919, 1920; BAUER and PETERS, 1920; FERRIER, 1922a, 1922b; EMDEN, 1920, 1922; v. GLEICH, 1928; CAMPBELL-TRUMPLER, 1923a, 1928). But it has been shown convincingly that in the light of our present knowledge such effects must be expected to be quite negligible. This holds also for the so-called Courvoisier Effect (COURVOISIER, 1920, 1932). Modern values for the refractive index in the solar atmosphere have been given, for instance, by PROISY (1949).

Other objections have been based on possible distortion of the photographic emulsion (SILBERSTEIN, 1920; SLOCUM, 1921; ROSS, 1920; see also GOLLNOW and HAGEMANN, 1956), and photometric effects (BOTTLINGER, 1920a, 1920b; WOLF, 1920). More serious are troubles possibly introduced in some circumstances by the

optical system (e.g. STROHMEYER, 1921). More recently also a possible deviation of photons in the gravitational field of the Sun has been discussed (WHEELON, 1952; PAPAPETROU, 1953).

Reports and discussions on the light-deflection have been frequently published ever since this topic was first raised by EINSTEIN. While the more specialized and important papers will be quoted individually in their proper context below, references to a number of more general papers are also given here. (EDDINGTON, 1918a, 1918b, 1918c; BOTTLINGER, 1920a, 1920b, FORSYTH, 1921a, 1921b; HEPPERGER, 1922; HOPMANN, 1928; TRUMPLER, 1929c; KOPFF, 1932; STOYKO, 1932; DANJON, 1932a, 1932b; HERMANN, 1936; DYSON and WOOLLEY, 1937; TIKHOV, 1937; BUCERIUS, 1938; OVENDEN, 1952; FREUNDLICH, 1952, 1953, 1955; TRUMPLER, 1956; MATTIG, 1956; MIKHAILOV, 1956, 1957, 1959. Some essentially historical remarks can be found in TRUMPLER, 1923; PERRINE, 1923; POOR, 1927; HALLUIN, 1942).

2. General Treatment of the Problem

Before giving a report of the actual observations and the results so far available, we would like to outline the general conditions which should be fulfilled to ensure successful observations and to point out the many difficulties the observers have to cope with.

The light deflection α, as predicted by Einstein's formula (2), can be represented by a hyperbola of the kind shown in Fig. 2. To decide whether the curve is a hyperbola

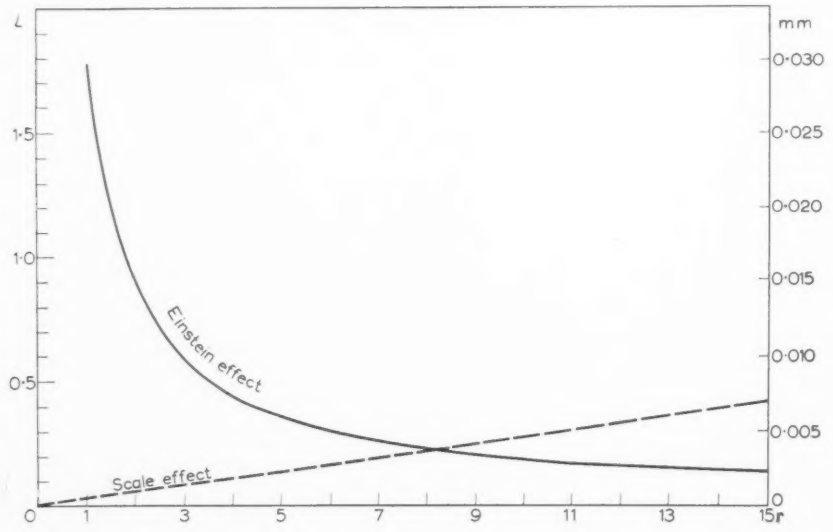


FIG. 2. This graph, a hyperbola, shows the behaviour of the predicted light-deflection, plotted as a function of the distance r from the centre of the Sun. The broken straight line indicates the "Scale Effect" (see p. 53), produced by an alteration of 0·1 mm in the focal setting of a "Normal Astrograph" (f = 343 cm).

Abscissae: distances from the centre of the Sun, expressed in units of the solar radius.
Left-hand ordinate: light-deflections in seconds of arc, as predicted by EINSTEIN (1916). See eq. (2).
Right-hand ordinate: the same, but expressed in millimetres on the photophraphic plate, assuming the focal length of the telescope to be 343 cm.

and, if so, to determine its form satisfactorily, is possible only if sufficient observations are available for $r < 2$. That is, enough stars must be available near the Sun's limb at the moment of totality, and their images on the photographic plate must be measurable with high precision. This condition raises the primary difficulty that the Sun passes rather few suitable stars in its yearly movement along the ecliptic, and that obviously such near and favourable approaches will coincide very seldom with the precise moment of a total eclipse. Furthermore, the Sun's corona becomes rapidly brighter when one approaches the Sun's limb, and therefore reduces seriously the contrast between the star images and their background on the photographic plate. This obliteration by the corona is the reason why at all eclipses which permitted successful observations of the light-deflection, the most interesting stars near the Sun's limb have been very difficult to measure, or were actually lost altogether. For the 1929 eclipse, which was observed through a sky of excellent transparency, some average values were determined of the limiting stellar magnitude which could be successfully observed near the Sun, as a function of their distance from the Sun's limb (FREUNDLICH, v. KLÜBER, and v. BRUNN, 1933a; v. KLÜBER, 1932b). A graph showing this relation is reproduced here in Fig. 3, but it must

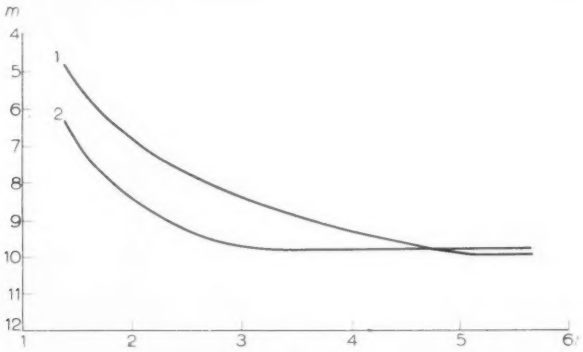


FIG. 3. This diagram indicates in a general manner the faintest stellar magnitude m which may on the average, be recorded on an eclipse photograph, taken with 60 seconds exposure time under favourable conditions, plotted as a function of the distance r of the star from the Sun's centre. (1) Telescope of 20 cm aperture and focal length $f = 343$ cm; (2) telescope of 20 cm aperture and $= 850$ cm. *Abscissae*: distances from the centre of the Sun in solar radaii. *Ordinates*: faintest stellar magnitude to be expected. (v. KLÜBER, 1932b).

be emphasized that only approximate values can be given in this way. The contrast between corona and star image depends, for instance, upon the colour temperature of the stars, upon the actual intensity of the corona at the position occupied by the star, and especially also upon the photographic emulsion and the processing (to which therefore much consideration has to be given). Taking into account that the spectral distribution of the corona light is identical with that of the Sun, it may perhaps be possible to reach somewhat fainter stars by using a suitable red filter and a red-sensitive emulsion. The intensity of the corona, furthermore, depends upon the cycle of solar activity, and quite appreciably upon the position angle with respect to the Sun's axis. At $r=2$, differences in the surface brightness of the corona of the order of more than one magnitude may occur (VAN DE HULST, 1950). Also larger f-ratios than those given in our example in Fig. 3 may help in reaching fainter starts. It follows from what we have just said that it will probably not be possible ever to measure

the full maximum value of 1″.75 of the deflection itself, which in other problems of astrometry would be a fairly large and easily measurable amount. The nearest stars so far observed were, with a single exception, at distances from the Sun's centre of about $r \geqq 2{\cdot}0$, corresponding to a deflection of the order of only 1″.0 (see below, Table 1, column 13). The most interesting part ($r < 2$) for the hyperbola in Fig. 2 is therefore hardly covered by observations at all.

Total solar eclipses are visible only from within a very narrow belt (100–200 kilometres) of a length of many thousands of kilometres, which the shadow of the Moon draws more or less at random over the surface of the Earth. Besides the trouble introduced by the light of the corona, the main and principal technical difficulty for observing the light-deflection is therefore, that it requires the absolute measurement of a very small quantity during a particular short time-interval under the usually quite difficult conditions of a temporary field-station in some more or less remote part of the world.

The technique of reducing plates for the determination of the light deflection is related to that used for many years in the photographic method for stellar parallaxes and proper motions (ZURHELLEN, 1904; KÖNIG, 1933). The principle is shown in a simple sketch in Fig. 4. The Sun with surrounding stars is indicated in (a) which

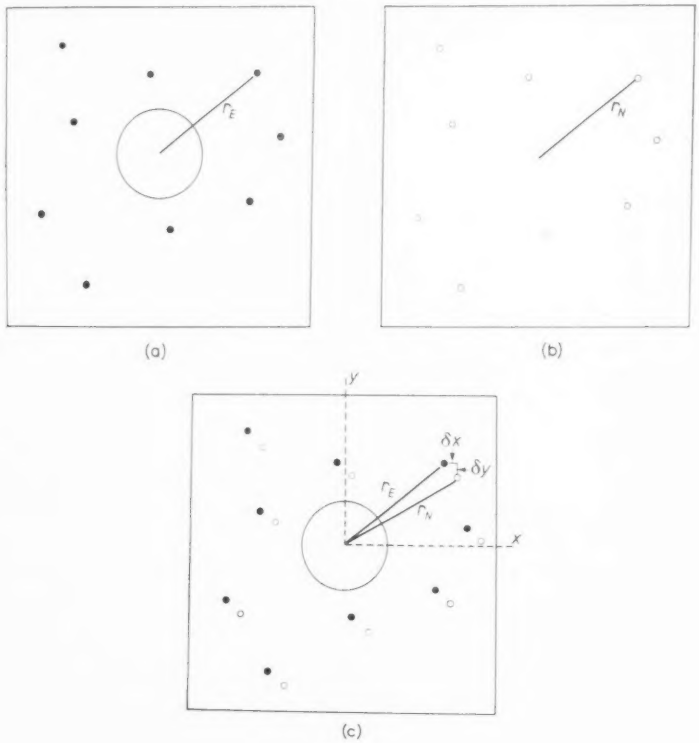


FIG. 4.

The notations are explained on p. 53.
(a) = "eclipse plate" showing the Sun and surrounding stars;
(b) = the corresponding "night plate" taken of the same star-field when visible at night, i.e. several months after (or before) the eclipse;
(c) = both plates combined, as they appear in the measuring machine, for the purpose of determining relative rectangular co-ordinate differences.

is usually called the eclipse-plate. The positions of the stars with the high accuracy needed here, cannot be obtained simply from a star catalogue. Therefore, the same star-field has to be photographed before the eclipse, or more often after, at a time when this field is visible in the night sky. This is done with the same instrument, and so far as possible when the field has the same position in the sky as during the eclipse. This produces the so-called night-plate (b) and, for convenience in later measurements, this is usually taken through the reverse side of the plate. For then the eclipse-plate and the night-plate can be adjusted emulsion to emulsion in the measuring-machine in such a way that very near to each star of the eclipse-plate the corresponding image from the night-plate will be visible (c). It is then easily possible to measure the relative position of each of such star-pairs in rectangular co-ordinates δx and δy as indicated in Fig. 4. Their values will depend upon the accidental way in which the two plates are clamped together in the measuring-machine (involving a lateral shift and a rotation); furthermore, on the possible change in the scale of the instrument (mainly due to the difference in the setting of the focus, which may have occurred between the exposure of the eclipse-plate and the exposure of the night-plate), on some other plate parameters, and on the light-deflection itself. Minor corrections may be necessary for differential atmospheric refraction, for differential aberration, for proper motion and parallax; but these can generally be applied in the usual way without much difficulty. If sufficient stars are available (at least 6), the various required parameters can be determined by a straightforward analytical method, and, finally, the Einstein effect L can be derived.

It will be seen from what follows that the main trouble in the reduction arises from the quite unavoidable fact that for instrumental reasons there will generally be a more or less conspicuous difference in the scale-value between the eclipse-plate and the night-plate. This can be easily visualized from our Fig. 2. Supposing the Sun is near the centre of the plate of a Normal Astrograph of focal length $f = 343$ cm; then the Einstein effect (full line) will shift the star's image at the Sun's limb by about $1''.75$, or by about 0·031 mm on the plate. At $r = 8$ solar radii from the centre of the Sun, the Einstein effect will have dropped to only $0''.27$ or about 0·005 mm. But if the effective focal length of the instrument or its focal setting (i.e. the distance between the principal plane of the objective and the surface of the photographic plate) has changed between the eclipse and the night exposure (which are necessarily separated by a period of at least 4 months) by only 1/10 mm, it would cause a scale-correction which is indicated in our sketch by the broken line, i.e. a deviation increasing linearly with distance from the plate's centre. At about $r = 8$ this scale-correction is already of the same order as the Einstein effect at the same distance. Whilst the Einstein effect is decreasing hyperbolically with distance from the Sun, the scale-correction is increasing linearly (Fig. 2). It is obvious that the determination of the Einstein effect will be more and more difficult with increasing distance of the available stars from the Sun, not only because it becomes much smaller, but mainly because the unavoidable scale error becomes more and more effective.

In this just outlined relative measurement of an eclipse-plate against the night-plate in rectangular co-ordinates each star i at a distance r_i from the Sun's centre will contribute one value δx_i and one value δy_i (Fig. 4). The analytical procedure to extract from these measurements the desired Einstein effect, and to separate it from

the other instrumental parameters, proceeds usually by forming for each star two equations of the form:

$$\delta x_i = A_x + B_x y_i + C_x x_i^2 + D_x x_i y_i + S_x x_i + L_x \frac{x_i}{r_i^2} + \\ + E_x x_i^3 + F_x x_i^2 y_i + G_x x_i y_i^2 + H_x y_i^2 + J_x y_i^3 \quad \ldots . (3)$$

$$\delta y_i = A_y + B_y x_i + C_y x_i y_i + D_y y_i^2 + S_y y_i + L_y \frac{y_i}{r_i^2} + \\ + E_y y_i^3 + F_y y_i^2 x_i + G_y y_i x_i^2 + H_y x_i^2 + J_y x_i^3, \quad \ldots . (4)$$

where we should have

$$B_x = -B_y, \quad C_x = C_y, \quad D_x = D_y, \quad E_x = E_y, \quad F_x = F_y, \quad G_x = G_y,$$
$$H_x = H_y, \quad J_x = J_y, \quad L_x = L_y.$$

To most of the coefficients a geometric significance can be given as follows:

A = arbitrary relative translatorial shift and
B = arbitrary relative rotational shift,
of the two plates during the measurements
C and D = inclination of the plate against the optical axis
S = Scale-value
L = light-deflection
E and F = scale-value, inclination, optical distortion
G = scale-value and inclination
H and J = bending of plate.

In most cases only the first six parameters of these equations have been used and these have often been found sufficient for solving the problem. For n stars we have $2n$ equations of this kind, and these are solved in the usual way by the method of least squares. The number of stars must obviously be at least equal to the number of unknown parameters used in the equations, i.e. usually at least six.

As we have already seen from purely geometric considerations there is a very unpleasant connection between the scale-value S and the Einstein effect L. This can be demonstrated, of course, even better by a detailed analytical investigation as it was carried out by several authors (e.g. FREUNDLICH and v. BRUNN, 1933; FREUNDLICH and GLEISSBERG, 1935; FREUNDLICH and LEDERMANN, 1944). Let us assume that we require a standard deviation σ_L for L of the order of $\pm$ 0".1; the weight W_L for L will then be $1/\sigma_L = 100$. From measurements of k plates the standard deviation σ_0 of δx_i and δy_i becomes

$$\sigma_0 = \pm \frac{0''.3}{\sqrt{k}},$$

or the weight

$$W_0 = \frac{k}{0{\cdot}09},$$

and therefore the corresponding relative weight

$$\frac{W_L}{W_0} \geqq \frac{9}{k}.$$

From the theory of least squares for the solution of normal equations *including* the parameters L and S, and for a fairly symmetrical star distribution of n stars, we get

$$\frac{W_L}{W_0} = n\left(\frac{1}{h} - \frac{1}{a}\right) \qquad \text{with}$$

$$\frac{1}{h} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{r_i^2} \qquad \text{and}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} r_i^2,$$

where r_i is the distance of the star i from the Sun's centre. The expression in the round bracket will always be a small quantity. If however the scale-value of S is *not* included in the equations, but is known independently by some suitably arranged observations, we would get the much more favourable expression

$$\frac{W_L}{W_0} = \frac{n}{h}.$$

The sensitive dependence of L on the scale-value S, and therefore the accuracy required for the determination of the scale value (or of the equivalent focal-lengths f), can be seen from the following expression, which can be derived in an elementary manner from the normal equations

$$\delta L = -h\delta S = -\bar{r}^2\delta S = -\frac{\bar{r}^2 \,.\, df}{f},$$

$\bar{r}$ being the mean distance of the stars from the centre of the Sun. If we want δL to be of the order of $0''.1$, i.e. 10^{-4} of the solar radius, and accept $r = 5$, we get

$$\frac{df}{f} = 4\,.\,10^{-6}$$

which means that the scale (in other words the effective focal-length) must be known to better than $\pm\ 10^{-5}$ (FREUNDLICH and v. BRUNN, 1933). This is a very strict experimental requirement in view of the conditions under which eclipse observations are made, for it means that the change of the focal setting for a Normal Astrograph (of $f = 343$ cm) between the exposure of the eclipse-plate and the night-plate should be known to less than $\pm\ 0{\cdot}03$ mm. Hence one of the main tasks of the observer is to determine the scale-value of his plates as accurately as possible. This, of course, has been realized right from the beginning and special observational techniques

have been developed to achieve this end. If no special arrangements for the determination of the scale-value are made, then the equations (3) and (4), containing the parameter S as one of the unknowns, must be applied straight away. As we have seen, the weight of the resulting L may be appreciably lowered by the coupling of S and L in the equations. To meet this difficulty, as early as 1919 observers had arranged to obtain photographs of an independent field of stars (the "check-field") at a sufficient distance from the Sun, with the same instrument, either during the eclipse itself or in the night before or after the eclipse; this procedure should provide some kind of check of the scale-value. A specially promising method was tried several times: during the eclipse itself the telescope was pointed first to the Sun, and then to a check-field so far away that the Einstein effect could be considered as negligible. The plates then carried two star fields superimposed upon each other, the intention being that the instrumental parameters of the equations and especially the scale-value would be obtained from the check-field, quite independent of the Einstein effect. A disadvantage of this procedure is that the movement of the telescope, even through a fairly small angle, may itself cause a small alteration of the focal setting which, as we remember, ought to be kept constant to about $\pm 0{\cdot}01$ mm (Freundlich, v. Klüber and v. Brunn, 1933). Probably it would be worthwhile to try the same procedure with a horizontal camera, merely swinging the coelostat mirror through a small angle during the eclipse (Freundlich, v. Klüber, and v. Brunn, 1931a).

In another method a plane-parallel plate (or a semi-transparent mirror, or even a real quartz-mirror which is smaller than the objective), tilted with respect to the optical axis of the telescope, has been used in front of the main objective (Mikhailov, 1949; van Biesbroeck, 1949; Popov and Fjodorov, 1954). The star-field surrounding the Sun is photographed through this plate, while at the same time a check-field at some distance from the Sun is reflected by the plate onto the same photographic plate. Again two star fields superimposed upon each other are obtained, in this case with strictly simultaneous exposures. This very attractive method has been criticized because the two star-fields are not imaged by strictly the same optical arrangement, the one being transmitted and the other one reflected by the semi-transparent plate. Any small distortion of this plate may affect the reflected beam, and thus the scale-value which is supplied by the check-field more than the transmitted beam (Freundlich, 1950). In principle, therefore, the scale-value taken from the check-field might again be not quite reliable. But if the dimensions of the reflecting mirror are kept small enough, so that each point of the check-field is imaged over the same whole mirror-surface, then the mirror will obviously not alter the scale-value of the check-field with respect to the eclipse-field by a significant amount; this method seems, therefore, to be rather promising. Its main disadvantage is that at the best it necessarily causes a loss of about half the light in each field.

Another very elaborate observing-method was developed and used at several eclipses by the Potsdam observers (Freundlich, 1930; Freundlich, v. Klüber and v. Brunn, 1931a, 1931b; v. Klüber, 1926, 1929a, 1929b, 1931; Mattig, 1956). They worked with a large horizontal double-camera, and the scale-value was again determined by an independent star-field during the eclipse itself. Furthermore, with the help of a large collimator, photographic reseaux were printed photographically on all plates to determine scale changes between the plates.

The scale value depends essentially on the distance between the objective and the emulsion of the photographic plate: SCHÜRER's (1954) attempt to control this distance by three invar wires was frustrated by poor weather.

Before considering in more detail the few expeditions so far successful in obtaining values for the Einstein effect, we should just mention briefly some other important technical requirements. The extremely high accuracy needed in controlling the effective focal setting obviously necessitates also very high accuracy in the definition of the position of the photographic plate, that is, of the plate holder on the telescope itself (KÖNIG and v. KLÜBER, 1941). Such techniques are known from parallax work.

Besides an independent determination of the scale value, as just explained, first-class guiding is absolutely essential; many observations in the past have failed in this respect. The scale-value of a Normal Astrograph ($f = 343$ cm) is $1'' = 0{\cdot}017$ mm, and star positions on the plates should be measured to about $\pm 0{\cdot}002$ mm, an accuracy not attainable unless the star images are symmetrical and of the best quality. Experience has shown that to guide rather large telescopes or coelostats under field conditions at an eclipse needs a great deal of consideration and special precautions (v. KLÜBER, 1932a).

Furthermore, high quality lenses and, if coelostats are used, quartz mirrors, are essential, despite the fact that usually a strictly differential method of reduction will be applied.

From general considerations a star-field with as symmetrical a distribution around the Sun as possible is desirable. As this depends entirely upon the accidental circumstances of the eclipse, this condition unfortunately is often not sufficiently fulfilled and then, if reductions are not carried out very carefully, systematic errors may result.

Because of their importance and the general interest of the results, nearly every published determination of the light deflection has led to a considerable amount of discussion. It is rather confusing to find that, in addition to the results published by the authors themselves, later discussions by others have often produced rather different figures from the same observational material. The chief reason for this is found in the fact already mentioned, that the final value for the light-deflection L is always quoted as extrapolated to the limb of the Sun, while the actual observations so far cover only ranges beyond $r > 2$, where the effect because of its hyperbolic decrease is very much smaller. A very small alteration in the much disputed scale-value or in the grouping or weighting of the stars, etc., can easily cause a rather large alteration of the result when extrapolated right up to the Sun's limb.

In Table 1 (column 16) we quote the numerical results in the same way as they were given by the authors themselves. Corrections proposed later by various critics will be mentioned below. There are, furthermore, two papers giving a more general discussion of the available measurements, one by DANJON (1932b, 1932c), and another by MIKHAILOV (1956, 1959). As these are both very interesting in demonstrating the nature of the relation between the observations and the final result, they are quoted below as well. In order to find possible corrections of the light-deflection and of the scale-value, after having eliminated the other parameters by a first approximation, DANJON proceeded in the following way: We denote by r_i the distance of the star from the centre of the Sun in units of the solar radius; by Δr_i the radial shift found for this star in the first approximation; by L the value of

Table 1

1	2 Observatory (site)	3 Eclipse	4 Local length f (cm)	5 Aperture ϕ (cm)	6 ϕ/f	7 Instrument (lens)	8 Field of plate	9 Number of plates	10 Exposure (sec)	11 Limiting stellar magnitude	12 Number of stars	13 r_{min} (solar radii from centre)	14 r_{max} (solar radii from centre)	15 Check-field	16 l.	17 m.s.e.	18 Reference
1	Greenwich (Brazil)	1919 May 29	570	10	1:57	Coelostat (Double)	2°4x2°0	7	28	6(ph)	7	2	6	no	1·98	0·16	Dyson-Eddington-Davidson, 1920
			343	20	1:17	Coelostat (Double)	2·7×2·7	16	5–10	6(ph)	11	2	6	no	0·93	—	
2	Greenwich (Principe)	1919 May 29	343	20	1:17	Coelostat (Double)	2·7×2·7	2	2–20	6(ph)	5	2	6	no	1·61	0·40	
3	Adelaide-Greenwich (Australia)	1922 Sept. 21	160	7·5	1:21	Astrograph (Quadruple)	7×8	2	20–30	8·3	11–14	2	10	yes	1·77	0·40	Dodwell-Davidson, 1924
4	Victoria (Australia)	1922 Sept. 21	330	15	1:22	Astrograph (Quadruple)	—	2	45	9·0	18	2	10	not used	1·75 1·42 2·16	—	Chant-Young, 1924.
5	Lick I (Australia)	1922 Sept. 21	450	12	1:37	Doubl. Astrograph (Double)	5×5	4	120–125	10·5(ph)	62–85	2·1	14·5	yes	1·72	0·15	Campbell-Trumpler, 1923a.
6	Lick II (Australia)	1922 Sept. 21	150	10	1:15	Astrograph (Quadruple)	15×15	6	60–102	10·4(ph)	145	2·1	42	yes	1·82	0·20	Campbell-Trumpler, 1928
7	Potsdam I (Sumatra)	1929 May 9	850	20	1:42	Coelostat (Double)	3×3	4	40–90	8·9	17–18	1·5	7·5	yes	2·24	0·10	Freundlich-v. Klüber-v. Brunn, 1931a.
8	Potsdam II (Sumatra)	1929 May 9	343	20	1:17	Astrograph (Triplet)	7·5×7·5	3	14–56	9·5	84–135	4	15	yes	—	—	Freundlich-v. Klüber-v. Brunn, 1933.
9	Sternberg (U.S.S.R.)	1936 June 19	600	15	1:40	Astrograph (Double)	3·5×3·5	2	25–35	9·6	16–29	2	7·2	not used	2·73	0·31	Mikhailov, 1949.
10	Sendai (Japan)	1936 June 19	500	20	1:25	Coelostat (Double)	2·9×2·9	2	80	8·6 (vis.)	8	4	7	no	2·13 1·28	1·15 2·67	Matukuma, 1940a.
11	Yerkes I (Brazil)	1947 May 20	609	15	1:40	Astrograph (Triplet)	4×4	1	185	10·2	51	3·3	10·2	not used	2·01	0·27	van Biesbroeck, 1949.
12	Yerkes II (Sudan)	1952 Feb. 25	609	15	1:40	Astrograph (Triplet)	4×4	2	60–90	8·6	9–11	2·1	8·6	yes	1·70	0·10	van Biesbroeck, 1953

VOL 3 1960

the light-deflection as predicted by EINSTEIN at the limb of the Sun; by k a possible correction factor to L; and by S a possible correction to the scale-value, as found in the first approximation. We then have:

$$\Delta r_i = \frac{L}{r_i} + S_i r_i, \qquad \text{or}$$

$$\frac{\Delta r_i}{r_i} = \frac{1{\cdot}75\ .\ k_i}{r_i^2} + S_i.$$

Inserting

$$x_i = \frac{1{\cdot}75}{r_i^2} \quad \text{and} \quad y_i = \frac{\Delta r_i}{r_i},$$

this can be written as the equation of a straight line (as shown as example in Fig. 5).

$$y_i = k_i x_i + S_i.$$

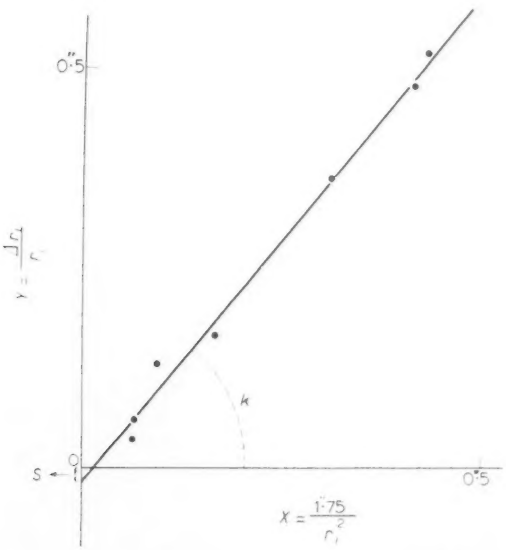


FIG. 5. Explanation of DANJON's method (pp. 58-59) for the final reduction of the measurements. The quantity $\tan k$ is the correction-factor to be applied to the predicted light-deflection L; S represents the scale-correction (DANJON, 1932c). The above figure illustrates the reduction of the Greenwich expedition of 1919, giving $\tan k = 1{\cdot}18$, and therefore $L = 1{\cdot}74 \times 1{\cdot}18 = 2{\cdot}06$; the scale-correction is $S = -0{\cdot}014$.

Its slope gives k, the correction to the light-deflection; and its intercept on the ordinate axis gives the correction to the scale-value. Both corrections can thus be visualized conveniently. This procedure again has been questioned because of the distribution of weights between the two unknowns (MIKHAILOV, 1956).

MIKHAILOV (1956), on the other hand, has taken most of the observations so far published and applied a final least-squares solution, using the equation of condition

$$\Delta r_i = \frac{L_i}{r_i} + S_i r_i$$

which gives final correcting values for L and for the scale parameter S. His results are quoted below.

3. SURVEY OF THE ECLIPSE RESULTS

After these more general considerations we give next a short account of the few expeditions so far successful in obtaining values for the Einstein effect. (See also Table 1.)

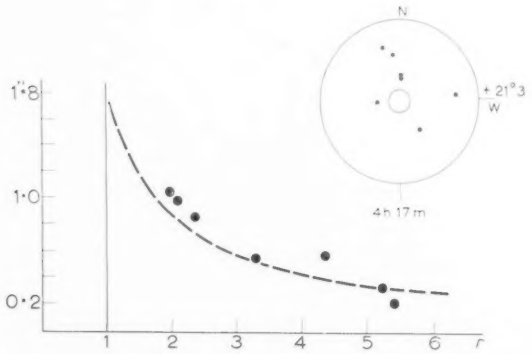
Fig. 6 represents all star-fields which have so far contributed to the determination of the light-deflection, up to a distance of about $8r$; they contain only those stars which have really been measured and used for the reductions. The same figure shows for each of these determinations the radial shift for each star (expressed in seconds of arc, as ordinates), plotted against its distance from the Sun's centre in units of the solar radius (i.e. 15 minutes of arc), as abscissae; only the values given by the authors themselves have been used. Different weights, as sometimes adopted by the authors, are not indicated. EINSTEIN's predicted deflection is shown in each

FIG. 6. These 9 combined diagrams show the actually measured light-deflections for each star, as far as available, using only the data given by the authors themselves, without having regard to individual weights or group-means. Some small amendments, mainly due to scale correction, may have to be applied to the one or the other of these observational sets. The broken hyperbola represents the Einstein Effect as it should be expected from theory.

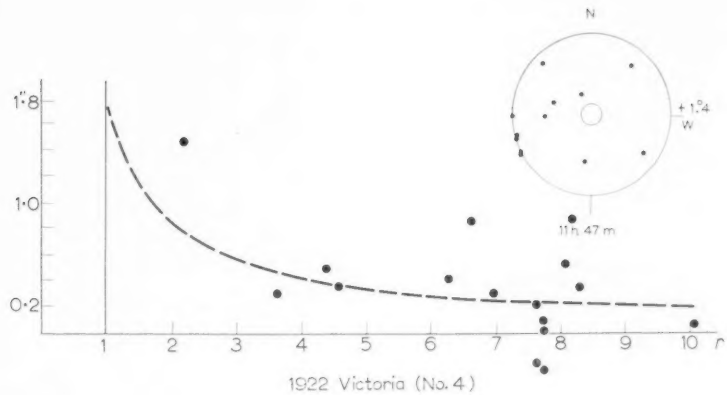
Abscissae: distances from the centre of the Sun in solar radii.

Ordinates: measured light-deflections in seconds of arc.

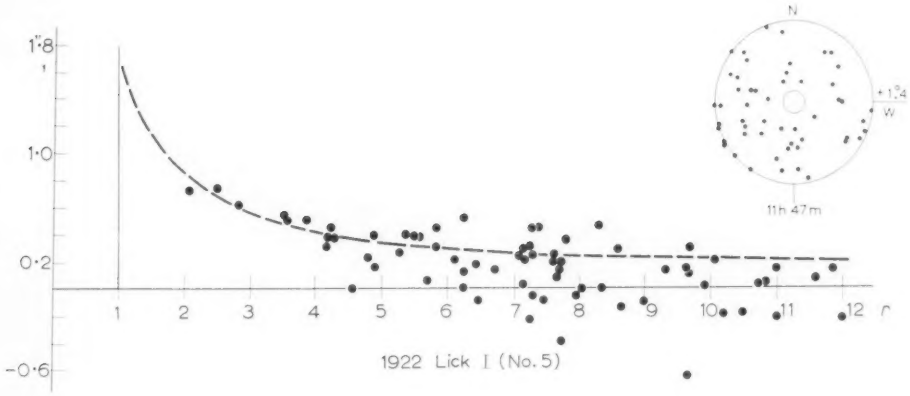
Inserted into the top right corners are the corresponding star fields, to a distance of about 8 solar radii from the Sun's centre. Only actually measured stars are plotted, without regard to the weight given by the observers. Co-ordinates are indicated, giving the positions of the Sun's centre for 1855·0 (B.D.-charts).



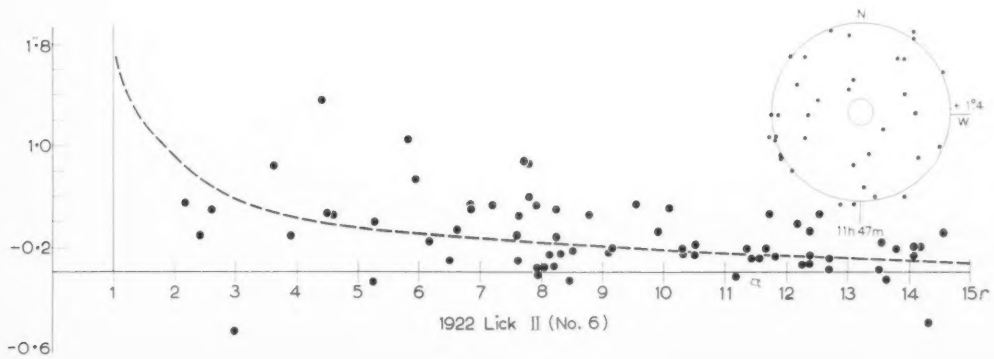
1919 Greenwich, (see Table 1, Nos. 1 and 2)



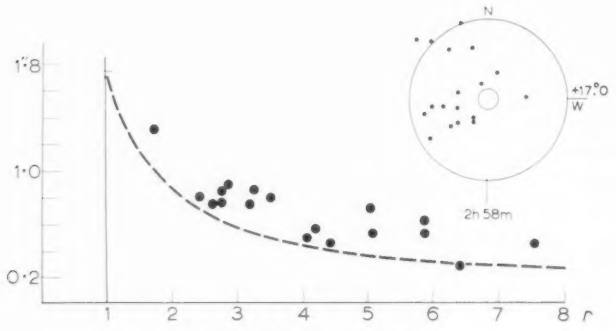
1922 Victoria (No. 4)



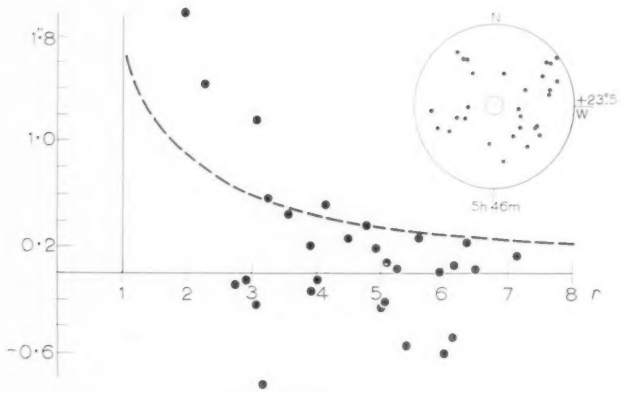
1922 Lick I (No. 5)



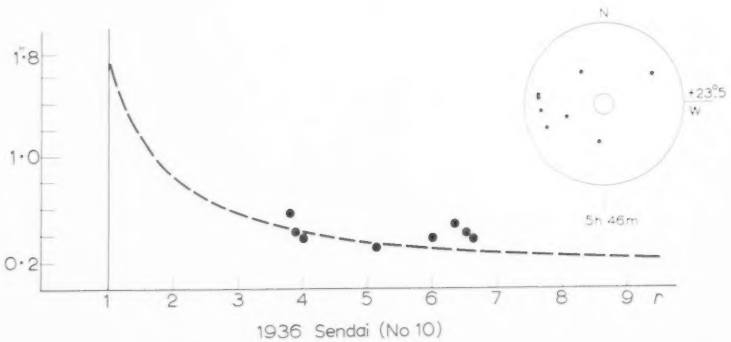
1922 Lick II (No. 6)



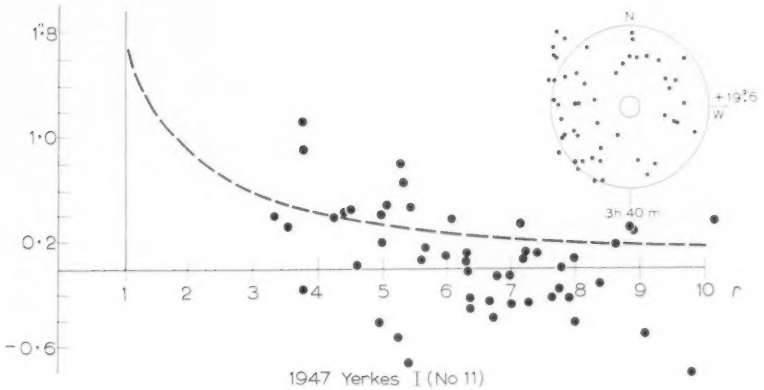
1929 Potsdam I. (No 7)



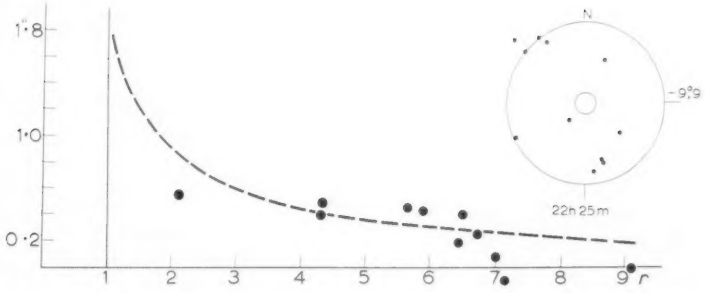
1936 Sternberg (No 9)



1936 Sendai (No 10)



1947 Yerkes I (No 11)



1952 Yerkes II (No. 12)

case by the dashed curves. These sketches emphasize the fact that the most important part of the hyperbola is scarcely covered by stars at all. In fact, most observations could be represented simply by straight lines (MIKHAILOV, 1956), and it is quite obvious that the actual deflection-law cannot be determined by the observations available at present. It is also necessary to keep in mind that in nearly every case represented in this figure, some small correction or other has been proposed by the authors or by various critics for reasons briefly mentioned above, mostly because of a correction in the scale-value, or because of different weighting or grouping of the stars or of the plates. In the following paragraphs the bold numbers are identical with those given in the first column of Table 1 and in the diagrams, (Fig. 6, etc.).

(**1**), (**2**) The first success in measuring the Einstein effect (EDDINGTON, 1919; Report joint Eclipse-Meeting, 1919; CROMMELIN, 1919a, 1919b; DAVIDSON and CROMMELIN, 1919; FREUNDLICH, 1920) was obtained by two British teams observing on 1919 May 29 from *Sobral* (Brazil) and from the Isle de *Principe* (Gulf of Guinea), respec-

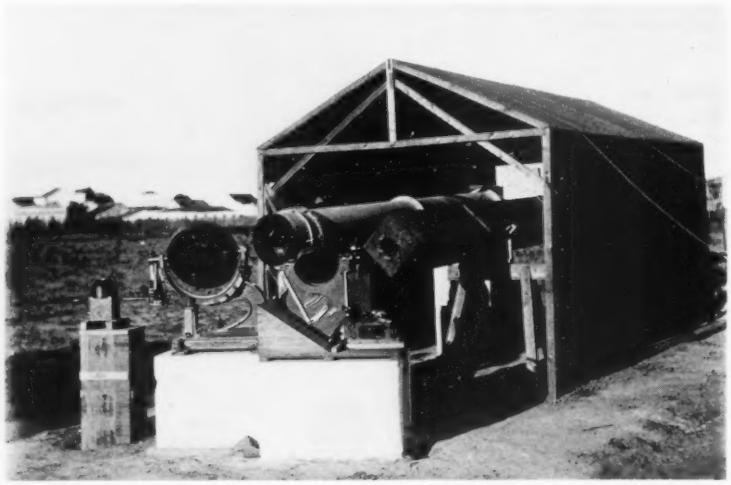


FIG. 7. The instrument used by the Greenwich expedition in 1919 at Sobral (Brazil). The two coelostats are feeding two horizontal telescopes; $f = 343$ cm and aperture 20 cm, on the left; $f = 570$ cm and 10 cm aperture, on the right; DYSON, EDDINGTON and DAVIDSON, 1920. (Photo. C. R. DAVIDSON.)

tively; (DYSON, EDDINGTON, DAVIDSON, 1920). A large number of plates were obtained with an astrograph and with horizontal telescopes fed by coelostats, but only a few of them were found suitable for reduction. No independent determination of the scale-value had been carried out, but for the observations on Principe photographs of a check-field were obtained. The result for the Sobral station, considered as the more accurate, is $L = 1''.98 \pm 0''.16$ m.e., and that for Principe is $L = 1''.61 \pm 0''.40$. The star-field for this eclipse is reproduced in Fig. 6, together with a graph giving the results for the individual stars. Fig. 7 shows the two horizontal telescopes used at Sobral ($f = 570$ and 343 cm) with their coelostats. Some discussion arose later from the fact that corrections for aberration and refraction had been incorporated in the least squares solution in such a way as actually to lower the

resulting weight, while all such corrections would better have been taken into account independently.

Objections were also raised because there were still some small tangential components left in the vectors for the radial shifts of the stars (RUSSELL, 1920), and because of suspected disturbances from refraction (BAUER, PETERS, 1920; ANDERSON, 1919–1924; DINES, RICHARDSON and ANDERSON, 1918; JEFFREYS, 1919; Royal Astronomical Society discussion, London, 1919a, 1919b; NEWALL, 1919; EDDINGTON and CROMMELIN, 1919), or because optical disturbances (STROHMEYER, 1921), or photographic ones (SLOCUM, 1921) were suspected. DYSON and WOOLLEY (1937) re-discussed these findings later, taking into account second-order terms in refraction and aberration, but no substantial alterations resulted. A re-discussion of the Sobral-values by HOPMANN (1923b), on the other hand, gave $L = 2''.16$.

Further re-discussions of the Sobral data by DANJON (1932) gave $L = 2''.06$, while MIKHAILOV (1956) obtained $L = 1''.95 \pm 0.088$ (or, with a scale-correction, $L = 2''.07 \pm 0.085$). DANJON's result is represented in Fig. 5.

(**3**) While on 1922 September 21 the weather was very unfavourable for expeditions working from Christmas Island (Indonesia), (FREUNDLICH, 1923a, HOPMANN, 1923a), a success was achieved at this eclipse by a combined team of Australian and British astronomers. The observations were made from *Condillo Downs, Australia* (DODWELL and DAVIDSON, 1924). This time a high-quality four-lens astrograph of the rather short focal length of 160 cm was used. A check-field was photographed during the eclipse and plate-parameters were taken from this check-field for the reduction of the star-field surrounding the Sun. The reduction of two plates gave $L = 2''.36$ and $L = 1''.18$, from which the observers deduced the mean value $L = 1''.77 \pm 0''.4$. The star field is very similar to that given in Fig. 6 for the Victoria team (No. 4), but the authors give no final values for the radial-shift of individual stars.

(**4**) At the same 1922 eclipse a Canadian team from the Dominion Astrophysical Observatory at Victoria also observed the light-deflection from *Wallal in Australia* (CHANT, 1923; YOUNG, 1923; CHANT and YOUNG, 1924). Two plates were obtained with a quadruple astrograph; reductions were made by using the linear expression of the equations only, and there was no independent determination of the plate parameters. The results were found to depend rather critically on the selection of stars used for the reduction. The authors gave the solution $L = 1''.75$ from all stars; but excluding various somewhat doubtful star-images, the resulting values for L became $1''.42$ or $2''.16$, with mean errors of about $\pm 0''.40$.

(**5**), (**6**) The same eclipse favoured by the fine weather at *Wallal, Australia*, as well as by a very good star-field, was also successfully observed by a Lick expedition at Wallal (CAMPBELL and TRUMPLER, 1923a, 1923b; CAMPBELL, 1923; FREUNDLICH, 1923b). These observations produced one of the best determinations of the light-deflection so far. Two instruments were used, a special double camera of $f = 450$ cm (Lick I), shown in Fig. 8; and a wide-angle quadruple-astrograph of $f = 150$ cm, (Lick II), (CAMPBELL and TRUMPLER, 1928). Each of the four plates of Lick I carried, superimposed on the eclipse-field, a check-field taken with the same instrument on the nights before and after the eclipse. Furthermore, other check-plates

from the same star-field were secured some months later. In the reductions the second-order terms were taken from the check-field and then, as a first step, the linear plate-parameters were determined from a large number of stars at $r > 2^\circ$ (without taking into account the Einstein effect which at such distances was considered negligible for this approximation). A last solution, and a small correction for the scale-value eventually gave as a mean for the four plates $L = 1''.72 + 0''.15$. Some small, remaining systematic deviations appeared indicated in the check-field; if a correction were applied for these, it would lead to small differences only, mainly for the stars more

FIG. 8. Double astrograph, parallactically mounted, ($f = 450$ cm, aperture 12 cm), as used by the Lick expedition (Lick I) in 1922 at Wallal, Australia, by CAMPBELL and TRUMPLER. (Photo: Lick Observatory.)

distant from the Sun; but the value for L at the Sun's limb would then go up to $L = 2''.05$. The results given by the authors were later disputed by the Potsdam observers (FREUNDLICH, v. KLÜBER and v. BRUNN, 1931a, 1931b, 1932b; TRUMPLER, 1932a, 1932b, 1932c), who, after some alterations in the method of reduction, derived from the same observations a value of $L = 2''.21$.

The application of DANJON's methods (1932c) to the observations of Lick I gave a value $L = 2''.00$ (or, with a small correction to the scale-value, $L = 2''.05$). MIKHAILOV's (1956) re-discussion gave $L = 1''.83 + 0''.20$, a figure very nearly identical

with the one given by the authors themselves. Here again it is very significant that a straight line would also fit the observations quite well, demonstrating that even this comparatively good material, with a good star-field and the largest number of stars used so far, is not sufficient to decide purely empirically the form of the deflection-law.

A report on the results of Lick I deals also with possible trouble from the COURVOISIER effect (HOPMANN, 1923b; TRUMPLER, 1924), whilst criticisms were raised because of too large a scatter of the individual values for the different stars (ESCLANGON, 1924a, 1924b, 1924c), or because there were still small tangential vectors left in the resulting radial shifts (PORTER, 1929; POOR, 1930; COMAS SOLÁ, 1928; TRÜMPLER, 1929c.) Even a quite different explanation of the deflection was proposed (GLEICH, 1931).

The short-focus instrument of Lick II (CAMPBELL and TRÜMPLER, 1928; TRÜMPLER, 1928), covering a field of $15^\circ \times 15^\circ$, gave plates which showed up to more than 500 stars with a limiting magnitude of about 10^m5; check-fields were taken in a similar manner as for Lick I. In the fairly elaborate reduction all third-order terms were included, and the scale values were first determined from the outer stars of the field. A second solution eventually gave a final and small scale-correction, and the Einstein effect became $L = 1''82 \pm 0''20$. A later reduction by the Potsdam observers deduced a larger value, $2''07$, from the same material.

The Lick II observations, too, have also been re-discussed; DANJON (1932c) used the more important stars only and was able to show how the presence and the weight given to a single star can influence the whole result. His value is $L = 2''07$ for the Lick II observations.

The radial shifts from the 15 best stars from both the Lick I and Lick II observations, are shown in Fig. 9 which gives a good idea as to how well on the whole the measured shifts agree with the expected light-deflection.

(**7**), (**8**) A further successful determination of the Einstein effect was made at the eclipse of 1929 May 9 in *Takengon* (Sumatra), after most careful preparation with elaborate instrumental equipment, by an expedition of the Astrophysical Observatory of Potsdam (FREUNDLICH, v. KLÜBER and v. BRUNN, 1931a, 1931b; FREUNDLICH, 1930; v. KLÜBER, 1929a, 1929b, 1931; WATTENBERG, 1932). The large instrument of Potsdam I was a double horizontal camera of $f = 850$ cm, so far the longest focal length used for this problem (Fig. 10). It was fed by a coelostat with a special drive of high precision, constructed by ZEISS in Jena (v. KLÜBER, 1932a), giving first-class automatic guiding, electromagnetically controlled by a chronometer. During the eclipse one of the cameras took exposures of the eclipse field, whilst the other one simultaneously photographed a suitable check-field, both using the same coelostat. Immediately before and again after the eclipse a scale reseau from a large thermally insulated collimator was photographed on all plates, using the coelostat again at the same reflection angles as during the eclipse. The same procedure was repeated about six months later with the same star field, then in the night sky. The reduction procedure demanded a very large amount of measurement and calculation, and was designed to give an independent absolute determination of the scale-value. For the collimator and the reseau the latter could be determined from the check-field and from this the scale-value of the eclipse-field could be found independently. That this whole cycle of observations could work very satisfactorily

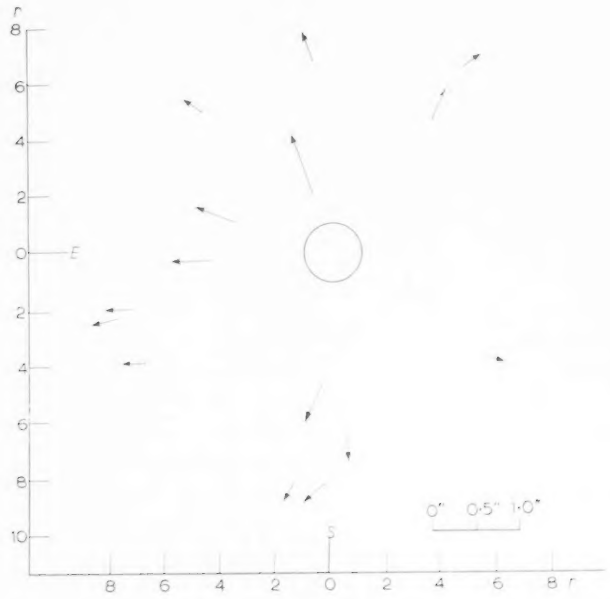


FIG. 9. Vector diagram of the radial shifts, derived from means of the 15 best stars by the Lick I and Lick II observations in 1922. It presents a very good indication of the existence of the light-deflection (CAMPBELL and TRUMPLER, 1928)

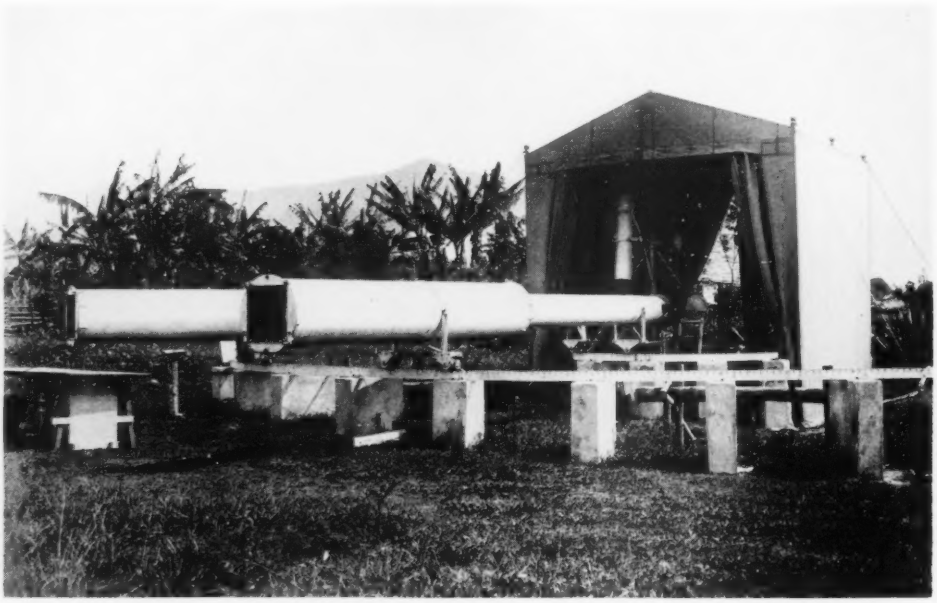


FIG. 10. Large double horizonta camera (f = 850 cm, aperture 20 cm) as used by the Potsdam expedition in 1929 to Takengon, Sumatra (Potsdam I). The coelostat and the collimator for printing the reseau are visible under the tent (FREUNDLICH, v. KLÜBER, v. BRUNN, 1931a). (Photo: v. KLÜBER.)

was demonstrated by a separate investigation, in which a star-field at night was treated by the same method as that employed at the eclipse itself (v. BRUNN and v. KLÜBER, 1937).

Favoured by very good weather, this instrument yielded four good plates; the outcome of the whole elaborate reduction was $L = 2''.24 \pm 0''.10$, showing a remarkably low m.e. The mean error of one plate was found to be $\pm 0''.30$. The star field at this eclipse and the radial shifts given by the different stars are shown in Fig. 6, while Fig. 11 represents the radial shift of all measured stars from the mean of all four plates.

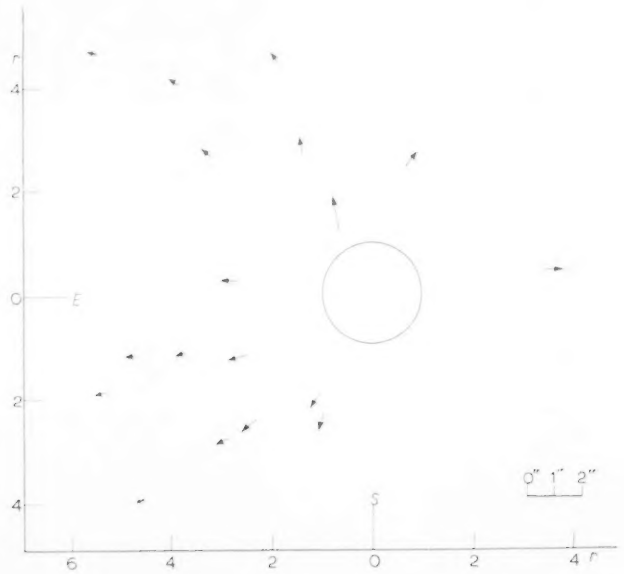


FIG. 11. Vector diagram indicating the observed light deflection for each star of the Potsdam I observations; mean of all four plates. The radial shift as expected by Einstein's theory is very clearly indicated (FREUNDLICH, v. KLÜBER, v. BRUNN, 1931a).

This paper again caused some controversy, partly because of the unfortunate asymmetry of the star field (LUDENDORFF, 1932; FREUNDLICH, v. KLÜBER and v. BRUNN, 1932b), and partly because small systematic residual errors were suspected for various reasons (TRUMPLER, 1932a, 1932b; FREUNDLICH, v. KLÜBER and v. BRUNN, 1932a, 1932b). LUDENDORFF deduced from the same material a value of $L = 1''.90 \pm 0''.15$, while TRUMPLER found that $L = 1''.75 \pm 0''.19$ satisfied the observations with good internal consistency. Another discussion by JACKSON (1931) produced the value $L = 1''.98 \pm 0''.20$. DANJON (1932c) found from his graphical method that he could represent the measures very well with $L = 2''.06$, while MIKHAILOV (1956), assuming a slight correction of the scale-value, gave $L = 1''.96 \pm 0''.11$.

Two further attempts, with the same instrument, at the eclipses of 1954 and 1955 failed because of bad weather (v. KLÜBER, 1955, 1956; MATTIG, 1956; NICHOLSON, 1956).

At the above mentioned eclipse of 1929 the Potsdam team used, in addition, a specially constructed strong astrograph, parallactically mounted, with a Zeiss triplet lens ($f = 343$ cm, $\phi = 20$ cm) of high quality (Potsdam II); see Fig. 12

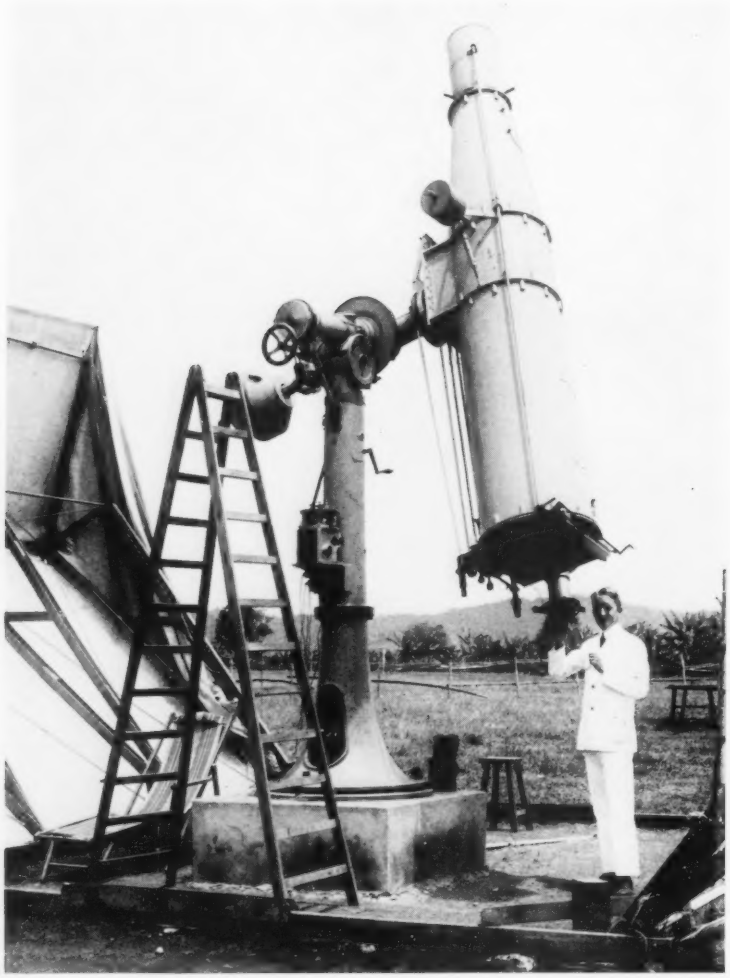


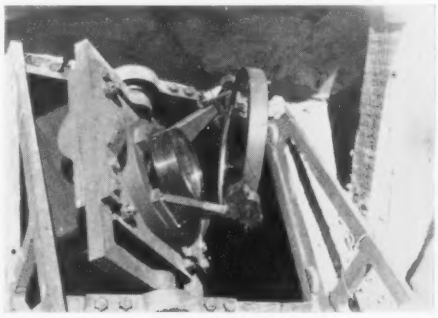
FIG. 12. Large parallactically mounted Zeiss astrograph ($f = 343$ cm, 20 cm aperture) with electrically controlled automatic drive, covering the large field of $7^{\circ}5 \times 7^{\circ}5$, as used in 1929 by the Potsdam observers (Potsdam II). During the eclipse itself a check star-field was photographed on each of the three plates taken of the Sun's surrounding, by pointing the astrograph alternately at a star-field distant from the Sun (FREUNDLICH, v. KLÜBER, v. BRUNN, 1933). (Photo: v. KLÜBER.)

(FREUNDLICH, v. KLÜBER and v. BRUNN, 1933). After each exposure of the star-field surrounding the Sun, the instrument was pointed during the eclipse itself onto a field at some distance from the Sun, so that each of the three plates obtained carried two star-fields superimposed on each other. Each of the plates, covering a field of $7^{\circ}5 \times 7^{\circ}5$, contained about 100 stars surrounding the Sun. This method seemed ideal for the determination of the plate parameters, and, especially, of the

scale-value. But even when all third-order terms were included in the reduction, and in spite of the fact that the star images were excellent, the resulting vectors for the different stars scattered widely. This unsatisfactory result was probably due to a very small relative mechanical deformation of the instrument, caused by moving the whole telescope during the eclipse from the one star-field to the other. Recalling what has been said above, namely, that the effective focal setting must be kept constant for both exposures to within a few hundredths of a millimeter, this failure is perhaps not surprising. The experiment with this instrument certainly demonstrates that this apparently promising method is in practice of rather doubtful value.

The extensive discussion of the Potsdam investigations by the observers themselves and by other authors has brought out very clearly, in an actual practical case, how critically the results depend upon the scale-value, how very important it is to obtain in some way or other an independent value for the scale-correction, and also how small the mechanical and optical tolerances of the instrument must be in order to secure good results (FREUNDLICH and v. BRUNN, 1933; KÖNIG, 1957).

(a)



(b)

FIG. 13

(a) A. A. MIKHAILOV's instrument, showing the specially designed plate-carrier and its drive, as used in 1936 and on later expeditions.

(b) MIKHAILOV's plane-parallel plate in front of the objective, for photographing simultaneously on the same plate the eclipse-field and a check-field.

(Photos: A. A. MIKHAILOV, Pulkovo Observatory.)

(9) At the eclipse of 1936 June 19, a Russian expedition from the Sternberg Observatory (MIKHAILOV, 1940a, 1940b, 1942, 1949) to *Kubishev* (Siberia), succeeded in obtaining further observations of the Einstein effect; (Fig. 13). This work included the first attempt to photograph the check-field and the eclipse-field simultaneously, by using a plane parallel plate in front of the objective of the telescope, as outlined above. Unfortunately, in taking the night photographs the check-field

was missed, so that the reductions had to be carried out in the usual straightforward manner by determining all the plate parameters from the eclipse field itself, using all second-order terms. Four different reductions of the two best plates gave $L = 2''.73 \pm 0''.31$, which is quite appreciably more than to be expected from EINSTEIN's theory. A later reduction (MIKHAILOV, 1956) gave $L = 2''.70 \pm 0''.40$. Another similar attempt by the Poltawa Observatory, in 1954, using a quartz mirror instead of the plane-parallel plate, failed because of bad weather (POPOV and FJODOROV, 1954).

(**10**) At the same eclipse of 1936 the Japanese Imperial University at Sendai made observations (SOTOME and HOSHIDA, 1936; MATUKUMA, 1940a, 1940b) at *Kosimizu* (Japan). A double-camera of $f = 500$ cm combined with a coelostat was used. Only one plate was obtained, and this showed not more than eight measurable stars. In the process of reduction this plate was combined with two different comparison night-plates; no special attempt was made to determine the scale-value separately; the measurements were reduced in the usual manner, including second-order terms, as well as a term for the scale-value, and another one for the light-deflection. The two plate-combinations gave $L = 2''.13 \pm 1''.15$ and $L = 1''.28 \pm 2''.67$, respectively, indicating excessively large mean errors. The diagram in Fig. 6 represents the mean of the two plates.

(**11**) Another attempt to use a plane-parallel plate in front of the objective of a long-focus astrograph ($f = 609$ cm), Fig. 14, was made by VAN BIESBROECK at the eclipse of 1947 May 20 from a station near *Bocajuva* (Brazil) (VAN BIESBROECK, 1949). A half-silvered plate was used, giving an equal loss of intensity in both the eclipse- and the check-field of about one magnitude. Only one photograph was obtained, on which the star images of the check-field appeared far worse than the images of the simultaneously exposed eclipse-field. In the opinion of the observer, this was probably caused by temperature distortion of the plane-parallel plate, and if so, this would be an interesting illustration of the objection mentioned above that the optical paths from the eclipse-field and the check-field are different. This leads to some doubt as to whether a scale-value determined from the check-field can safely be used for the reduction of the eclipse-field (FREUNDLICH, 1950; VAN BIESBROECK, 1950). In this case the check-field was not used. A simple graphical reduction like the one suggested by DANJON (1932c) was used by the author and gave $L = 2''.01 \pm 0''.27$. MIKHAILOV (1956) raised some objections to this simplified reduction, and in a re-reduction he found $L = 2''.20 \pm 0''.35$.

(**12**) At the eclipse of 1952 February 25, observing at *Khartoum* (Sudan), VAN BIESBROECK successfully obtained two more plates employing the same method as 1947, with special precautions to protect the plane-parallel plate from any kind of optical distortion. In a short paper (VAN BIESBROECK, 1953) the author derives from these two plates the value $L = 1''.70 \pm 0''.10$; the mean error is remarkably low. Unfortunately all stars, except one at $r = 2 \cdot 1$, are at a distance from the Sun $> 4 \cdot 3$, so that the most vital part of the expected hyperbola is not covered by stars at all. The figures given in this publication have also been re-reduced by MIKHAILOV (1956), who then obtained $L = 1''.43 \pm 0''.16$.

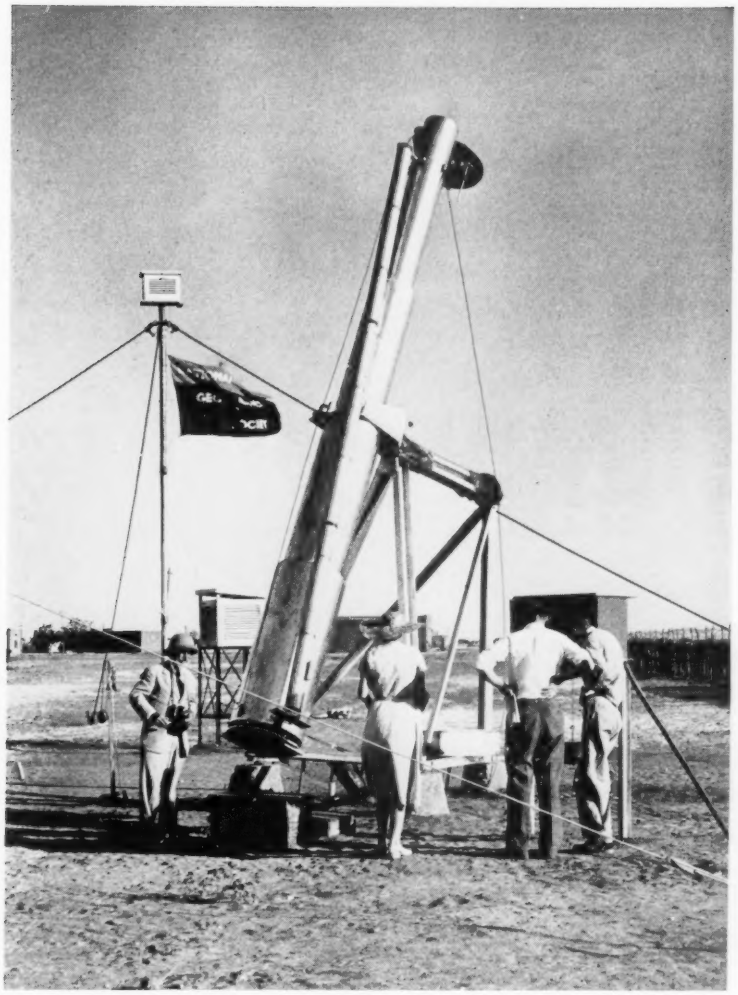


FIG. 14. Telescope (f = 609 cm, aperture 15 cm) with special mounting and drive, as used for the Yerkes II observations in 1952. A check-field was superimposed on the eclipse field during the eclipse itself, using a semi-transparent mirror in front of the objective; (VAN BIESBROECK, 1953). (Photo: v. KLÜBER.)

4. CONCLUSIONS

Summing up the results of the foregoing investigations, which comprise all observations available so far (1959), the following points can be made:

(a) Without exception, all observations indicate clearly that a light-deflection effect of the kind expected quite obviously exists in the neighbourhood of the Sun. But the observations are not sufficient to show decisively whether the deflection really follows the hyperbolic law predicted by the General Theory of Relativity, mainly because so far it has not been possible to obtain a satisfactory number of

star-images sufficiently near to the Sun. As things are at present, most observations could be represented quite well even by straight lines (MIKHAILOV, 1956).

(b) If the existing observations are extrapolated to the limb of the Sun, according to the supposed hyperbolic law, then there is an indication that the resulting constant L appears to be somewhat larger than the value expected by the General Theory of Relativity. Whether this behaviour is real, or only introduced by some systematic observational error, cannot at present be decided with certainty; (PAPAPETROU, 1953). WHEELON (1952) recently pointed out that if one were to assume photons with a non-vanishing rest-mass, then the effect of the non-zero mass would increase the light-deflection in accordance with the values for the rest-mass and the light-frequency. But if larger effects were produced in this way, there should also exist certain laboratory phenomena which are not known to occur.

(c) The fairly large discrepancies between the results of various authors, even when based on the same observational material, is not so surprising if one recalls that the value for L is in fact an extrapolation which is very sensitive to small alterations in the position of the actually observed points, if these lie rather far out on the supposed hyperbola. Furthermore, the quantity sought is unfortunately strongly coupled to the scale-value itself, which is very difficult to determine. A good impression of all available results may be obtained from the set of graphs in Fig. 6 and from Table 1.

To determine the relativistic light-deflection as accurately as possible is without any doubt a most important experiment because of its relation to fundamental consequences of the Theory of Relativity. As observations of this kind are so difficult and can be carried out only at total solar eclipses, progress in future will probably be rather slow. From the experience gained so far, the following points seem to be important for further investigations:

(I) First of all, every possible effort should be made to obtain somehow or other an independent determination of the scale-value. This is probably an absolutely essential condition for achieving any real progress. If any other plate-parameter could be determined independently as well, it would certainly increase the weight of the result considerably.

(II) The usual minimum number of stars needed in the eclipse field is six, but actually as many stars as possible are required; and if no stars are available for $r < 2$, the actual law of deflection (supposed to be hyperbolic) can probably not be determined by the observations alone. Furthermore, a symmetrical distribution of the stars around the Sun is most desirable. As the limit for measurable star images will be probably round about the tenth photographic magnitude, all these conditions place a rather severe restriction on the star fields suitable for further observations.

Certainly not every future total eclipse will fulfil these conditions. Possibly more star images could be obtained through the corona, by working in the red part of the spectrum only and/or by making use of the radial polarization of the corona light.

(III) First-class guiding is another essential condition, and we must emphasize most strongly how difficult this is to obtain at the temporary field-station of an eclipse expedition. Many observers in the past have underestimated this difficulty. The focal

length of the instrument should be larger than 300 cm, if possible at least 600 cm; a suitable aperture ratio would be about 1 : 40. A larger value may cause some loss of star images because of the larger apparent surface brightness of the corona. The whole instrument must be constructed so rigidly that under working conditions the scale-value can be maintained to something of the order of 10^{-6}. This in itself is a very severe condition.

(IV) High-quality optics are desirable, but it must be kept in mind that optics which are too complicated may, under field conditions, possibly not fulfil some of the requirements just mentioned. Mirrors, of course, should be made of fused quartz.

(V) It is in principle desirable to obtain several plates at an eclipse, since this increases the weight of the results appreciably.

Finally, we should like to state quite clearly that further observations of the light-deflection are only justified, if real progress is to be expected as a result of fulfilling as nearly as possible the stringent conditions summarized above. Even so, such observations remain among the most difficult of all those which can be attempted at a total solar eclipse.

REFERENCES

ANDERSON, A.	1919	*Nature (Lond.)*, **104**, 354.
	1920a	*Nature (Lond.)*, **104**, 436.
	1920b	*Nature (Lond.)*, **104**, 563.
	1922	*Astron. Nachr.*, **215**, 197.
	1924	*Astron. Nachr.*, **220**, 205.
BAUER, L. A. and PETERS, W. J.	1920	*Phys. Rev.* (2), **15**, 527.
BOTTLINGER, K. F.	1920a	*Jahrbuch der Radioaktivität und Elektronik*, **17**, 146.
	1920b	*Astron. Nachr.*, **211**, 239.
BRUNN, A. v. and KLÜBER, H. v.	1937	*Z. Astrophys.*, **14**, 242.
BUCERIUS, H.	1938	*Astron. Nachr.*, **263**, 153.
CAMPBELL, W. W.	1919	*Observatory*, **42**, 298.
	1923	*Publ. Astron. Soc. Pacific*, **35**, 11.
CAMPBELL, W. W. and TRUMPLER, R.	1923a	*Lick Obs. Bull.*, **11**, 41.
	1923b	*Publ. Astron. Soc. Pacific*, **35**, 158.
CAMPBELL, W. W. and TRUMPLER, R. J.	1928	*Licks Obs. Bull.*, **13**, 130.
CHANT, C. A.	1923	*Sciences C.N.S.I.*, **57**, 469.
CHANT, C. A. and YOUNG, R. K.	1924	*Publ. Dominion Astr. Obs.*, **2**, 275.
COMAS SOLÁ, J.	1928	*Rev. Soc. Astron. Esp. y Amer.*, **18**, 65.
COURVOISIER, L.	1920	*Astron. Nachr.*, **211**, 305.
	1932	*Astron. Nachr.*, **244**, 279.
CROMMELIN, A. C. D.	1919a	*Nature (Lond.)*, **104**, 280.
	1919b	*Astron. Nachr.*, **210**, 79.
CURTIS, H. D.	1913	*Publ. Astron. Soc. Pacific*, **25**, 77
DANJON, A.	1932a	*J. Phys. Radium*, (7), **3**, 15.
	1932b	*C.R. Acad. Sci. (Paris)*, **194**, 252.
	1932c	*J. Phys. Radium*, (7), **3**, 281.
DAVIDSON, C. and CROMMELIN, A. C. D.	1919	*J. Brit. Astron. Assoc.*, **29**, 45.
DINES, W. H., RICHARDSON, L. F. and ANDERSON, A.	1919	*Nature (Lond.)*, **104**, 393.
DODWELL, G. F. and DAVIDSON, C. R.	1924	*M.N.*, **84**, 150.
DYSON, F. W., EDDINGTON, A. S. and DAVIDSON, C.	1920	*Phil. Trans.*, **220A**, 291. (See also Smithsonian Report for 1919, **133**, 1921).
DYSON, F. and WOOLLEY, R. v. d. R.	1937	*Eclipses of the Sun and Moon*, p. 43. Oxford Univ. Press.

Eddington, A. S.	1918a	*Nature (Lond.)*, **101**, 15.
	1918b	*Nature (Lond.)*, **101**, 34.
	1918c	*Nature (Lond.)*, **101**, 126.
	1919	*Observatory*, **42**, 119.
Eddington, A. S. and Crommelin, A. C. D.	1919	*Nature (Lond.)*, **104**, 372.
Einstein, A.	1908	*Jahrbuch der Radioaktivität und Elektronik*, **4**, 411.
	1911	*Ann. Phys. (Leipzig)*, **35**, 898.
	1916	*Ann. Phys. (Leipzig)*, **49**, 769.
Emden, R.	1920	*Ber. Bayr. Akad. Wiss. München Math.-phys. Klasse*, p. 387.
	1922	*Astron. Nachr.*, **215**, 397.
Esclangon, E.	1924a	*J. Phys. Radium* (6), **5**, 10.
	1924b	*Lyon Obs. Bull.* **6**, 31.
	1924c	*C.R. Acad. Sci. (Paris)*, **178**, 196.
Ferrier, M.	1922a	*C.R. Acad. Sci. (Paris)*, May, 29.
	1922b	*Nature (Lond.)*, **109**, 831.
Forsyth, A. R.	1921a	*M.N.*, **82**, 2.
	1921b	*Observatory*, **44**, 355.
Freundlich, E. F.	1913	*Astron. Nachr.*, **193**, 369.
	1914	*Astron. Nachr.*, **197**, 335.
	1920	*Naturwissenschaften*, **8**, 667.
	1923a	*Astron. Nachr.*, **218**, 13.
	1923b	*Naturwissenschaften*, **11**, 962.
	1930	*Naturwissenschaften*, **18**, 313.
	1950	*Astron. J.*, **55**, 245.
	1952	*Wissenschaftliche Ann. Berlin*, **1**, 35.
	1953	*Phys. Blätt.*, **9**, 14.
	1955	*Vistas in Astronomy*, (Ed. A. Beer), Vol. 1, 239. Pergamon Press, London.
Freundlich, E. and Brunn, A. v.	1933	*Z. Astrophys.*, **6**, 218.
Freundlich, E. and Gleissberg, W.	1935	*Publ. Istanbul Univ. Obs., N.F.*, **1**, 44.
Freundlich, E. F., Klüber, H. v. and Brunn, A. v.	1931a	*Abhandl. Preussische Akad. Wiss. Berlin, Phys. Mat. Klasse*, No. 1.
	1931b	*Z. Astrophys.*, **3**, 171.
	1932a	*Astron. Nachr.*, **244**, 415.
	1932b	*Z. Astrophys.*, **4**, 221.
	1933	*Ann. Bosscha Sterrewacht Lembang (Java)*, **5**, 2.
Freundlich, E. F. and Ledermann, W.	1944	*M.N.*, **104**, 40.
Gleich, G. v.	1928	*Z. Phys.*, **51**, 740.
	1931	*Astron. Nachr.*, **243**, 133.
Gollnow, H. and Hagemann, G.	1956	*Astron. J.*, **61**, 399.
Halluin, H. D'.	1942	*Ciel et Terre*, **58**, 1.
Hepperger, J.	1922	*Astron. Nachr.*, **216**, 321.
Hermann, H.	1936	*Z. Phys.*, **100**, 667.
Hopmann, J.	1923a	*Himmelswelt*, **33**, 31.
	1923b	*Phys. Z.*, **24**, 476.
	1929	*Handbuch d. Physik*, **21**, 683.
Hulst, H. C. van de	1950	*Bull. Astron. Inst. Netherlands*, **11**, 135.
Jackson, J.	1931	*Observatory*, **54**, 292.
Jeffreys, H. J.	1919	*M.N.*, **80**, 138.
Jonckheere, R.	1918	*Observatory*, **41**, 215.
Klüber, H. v.	1926	*Die Sterne*, **1926**, 157.
	1929a	*Die Sterne*, **9**, 85.
	1929b	*Die Sterne*, **9**, 125.
	1931	*Die Sterne*, **11**, 141.
	1932a	*Z. Instrumentenkunde*, **52**, 381.
	1932b	*Astron. Nachr.*, **244**, 307.
	1955	*Naturwissenschaften*, **42**, 3.
	1956	*Naturwissenschaften*, **43**, 28.
König, A.	1933	*Handbuch d. Astrophys.*, Vol. I, p. 502, Springer, Berlin.
	1957	*Z. Astrophys.*, **42**, 87.
König, A. and Klüber, H. v.	1941	*Z. Astrophys.*, **20**, 311.
Kopff, A.	1932	*Naturwissenschaften*, **20**, 486.
Laue, M. v.	1920	*Phys. Z.*, **21**, 659.
Lewis, J. M.	1931	*Publ. Amer. Astron. Soc.*, **6**, 265.
Lindemann, A. F. and F. A.	1916a	*M.N.*, **77**, 140.
	1916b	*Observatory*, **40**, 47.
Lindemann, F. A.	1918	*Observatory*, **41**, 323.
Ludendorff, H.	1932	*Astron. Nachr.*, **244**, 321, 415.

MATTIG, W.	1956	*Die Sterne*, **32**, 177.
MATUKUMA, T.	1940a	*Japan J. Astron. and Geophys.*, **18**, 51.
	1940b	*Nature (Lond.)*, **146**, 264.
MIKHAILOV, A. A.	1940a	*C.R. Acad. Sci. USSR (N.S.)*, **29**, 189.
	1940b	*Phys. Ber.*, **22**, 934.
	1942	*Astron. J. USSR*, **19**, 1.
	1949	Report by V. G. FESENKOV: Expeditions to observe the total solar eclipse of 1941 September 21; *Publ. Acad. Science USSR*, p. 337, 1949.
	1956	*Astron. J. USSR*, **33**, 912.
	1957	*Acta Astronomica Sinica*, Vol. 5, 167.
	1959	*George Darwin Lecture*, Roy. Astron. Soc. 1958. M.N. **119**.
NEWALL, H. F.	1919	*M.N.*, **80**, 22.
	1920	*Observatory*, **43**, 145.
NICHOLSON, T. D.	1956	*The Scientific Monthly*, **82**, 221.
OVENDEN, M. W.	1952	*Sci. Progr.* **40**, 645.
PAPAPETROU, A.	1953	*Fortschritte der Physik*, **1**, 29.
PERRINE, C. D.	1923	*Astron. Nachr.*, **219**, 281.
POOR, CH. L.	1927	*J. Roy. Astron. Soc. Canada*, **21**, 225.
	1930	*J. Opt. Soc. Amer.*, **20**, 173.
POPOV, N. A. and FJODOROV, E. P.	1954	*Astron. Circ. USSR*, No. 151, 21.
PORTER, J. C.	1929	*Publ. Astron. Soc. Pacific*, **41**, 171.
PROISY, P.	1949	*Ann. Astrophys.*, **12**, 123.
Report Joint Eclipse-Meeting	1919	*Observatory*, **42**, 389.
ROSS, F.	1920	*Astrophys. J.*, **52**, 98.
Royal Astronomical Society, London.	1919a	*M.N.*, **80**, 96. (Discussion on the Theory of Relativity).
	1919b	*Observatory*, **43**, 33; (as 1919a).
RUSSELL, H. N.	1920	*M.N.*, **81**, 159.
SCHÜRER, M.	1954	*Orion*, Schaffhausen, No. 45, p. 309.
SILBERSTEIN, L.	1920	*M.N.*, **80**, 630.
SLOCUM, F.	1921	*Popul. Astron.*, **29**, 273.
SOLDNER, J.	1801	*Astron. Jahrbuch f.d. Jahr.* 1804, p. 161, Berlin; See also the article in:
	1921	*Wiedemann Ann.* (4), **65**, 593.
SOTOME, K. and HOSHIDA, G.	1936	*Tokyo Astron. Obs. Rep.*, **4**, 137.
STOYKO, N.	1932	*Bull. Astron. (Paris)*, **7**, 313.
STROHMEYER, C. E.	1921	*Nature (Lond.)*, **107**, 682.
TIKHOV, G. A.	1937	*C.R. Acad. Sci. Moscow (N.S.)*, **16**, 199.
TRUMPLER, R. J.	1923	*Publ. Astron. Soc. Pacific*, **35**, 185.
	1924	*Publ. Astron. Soc. Pacific*, **36**, 221.
	1928	*Publ. Astron. Soc. Pacific*, **40**, 130.
	1929a	*Publ. Astron. Soc. Pacific*, **41**, 23.
	1929b	*J. Roy. Astron. Soc. Canada*, **23**, 208.
	1929c	*Popul. Astron.*, **37**, 355.
	1932a	*Z. Astrophys.*, **4**, 208.
	1932b	*Publ. Astron. Soc. Pacific*, **44**, 167.
	1932c	*Science*, **75**, 538.
	1956	*Helv. Phys. Acta, Basel*, Suppl. IV, **106**.
VAN BIESBROECK, G.	1949	*Astron. J.*, **55**, 49.
	1950	*Astron. J.*, **55**, 247.
	1953	*Astron. J.*, **58**, **87**.
WATTENBERG, D.	1932	*Weltall*, **31**, 141.
WHEELON, A. D.	1952	*Phys. Rev.* (2), **85**, 383.
WOLF, M.	1920	*Astron. Nachr.*, **212**, 181.
YOUNG, R. K.	1923	*J. Roy. Astron. Soc. Canada*, **17**, 129.
ZURHELLEN, W.	1904	*Darlegung und Kritik der zur Reduktion photographischer Himmelsaufnahmen aufgestellten Formeln und Methoden;* Dissertation, Bonn.

The Clock Paradox

W. Cochran

Cavendish Laboratory, Cambridge, England.

Summary

Let us suppose that two clocks of identical construction are found to run in perfect synchronism while they are side by side, and that this synchronism is maintained even if one of the clocks, while remaining very near the other, is subjected to accelerations, however great. The clocks are then widely separated from one another with a relative velocity comparable with the velocity of light, and subsequently reunited. Are their readings necessarily identical? This question has caused controversy at intervals during the last fifty years. The controversy has been renewed recently when the apparently imminent advent of space travel has brought to the attention of a wider audience some of the striking predictions of the theory of relativity, concerning the possibility of "asymmetric ageing". In this article the correctness of the result deduced by Einstein in 1905, that the two clocks will in general *not* agree, is defended from a number of points of view, including the experimental evidence. The effect of a permanent gravitational field on the clocks is also discussed, and questions concerning the rate of a clock in an Earth satellite and the ageing of space travellers in a gravitational field are shown to have unexpected and amusing answers.

In the early years of this century there were many scientists who rejected Einstein's theory of relativity as being contrary to all reason. At the present time there are a few (there may be many, but only a few have expressed the view in print) who insist that one of the predictions in Einstein's first paper on the subject must be erroneous, for if correct the result would be contrary to the principle of relativity! The discussion of this particular point has recently been revived in the form of the question whether a space traveller from the Earth, whose velocity relative to the Earth over a considerable part of his journey was comparable with the velocity of light, would find on his return that the duration of the journey was less by his clock than by an identical clock which had remained on the Earth, or in other words that he, the traveller, had aged less than his stay-at-home twin. The discussion has aroused a great deal of popular interest—but then, it is not often that pure physics touches so directly on questions of human mortality.

The problem is discussed in this article in non-mathematical language, and from the point of view of an experimentalist.

1. Einstein's Paper of 1905

The equations of the special Lorentz transformation were derived by Einstein in this famous paper. They are

$$x' = \beta(x - vt),\ y' = y,\ z' = z \text{ and } t' = \beta(t - vx/c^2),$$

where $\beta = (1 - v^2/c^2)^{-\frac{1}{2}}$. They relate the co-ordinates (x, y, z) and time t of an event as measured in a frame of reference K to the co-ordinates (x', y', z') and time t' of the same event measured in a frame K' which relative to K has a uniform velocity v in the positive direction of x. (Fig. 1.) In the limit $v \ll c$ with which we ordinarily have to deal,

$$x' = x - vt,\ y' = y,\ z' = z \text{ and } t' = t,$$

the common-sense result of pre-relativity kinematics, which in particular asserts that time is absolute and independent of the frame of reference. Relativity physics abandons the idea of absolute time completely.

To illustrate this last point, let us suppose that we have three clocks A, B and C of identical construction which remain in perfect synchronism when all are together. Suppose now that A and B are widely separated and at rest in the frame K, with A at the origin and B at distance X along the x-axis. Since they are widely separated, A and B must be synchronized by an exchange of light signals, in accordance with EINSTEIN's specification, which is that if a signal is sent out from A at time t_1 and received back at time t_2 after reflection from B, then for synchronization to be achieved, B must read $\frac{1}{2}(t_1+t_2)$ at the instant of reflection of the light. The third clock C is supposed to remain throughout at the origin of the frame K', it coincides with A just as A reads $t = 0$ and C is synchronized with A by reading $t' = 0$ at this coincidence. After a time $T = X/v$, (measured in K) clock B will read $t = T$ and C

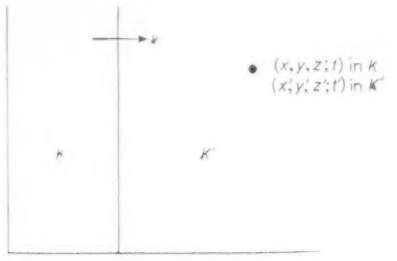


Fig. 1.

will be passing B. At this coincidence C will record, not $t' = T$ but $t' = \beta(T - vX/c^2)$, or $t' = T/\beta$ since $X = vT$. On meeting, therefore, C is behind B by an amount $T(1-1/\beta) \simeq \frac{1}{2}Tv^2/c^2$. This is the result which is often stated rather loosely as "a moving clock runs slow". We shall subsequently refer to the above thought-experiment as *Situation 1*, or S.1 for short. There is of course no paradox here, despite the fact that we might equally well regard the frame K' as being at rest, and K as moving, for any one clock in K will also run slow if it is compared with two successive clocks at rest in K'. There is no lack of reciprocity between the two frames in S.1.

We can best take the next step by quoting EINSTEIN's own words: "From this there ensues the following peculiar circumstance. If at the points A and B of K there are stationary clocks which, viewed in the stationary system, are synchronous; and if the clock at A is moved with velocity v along the line AB to B, then on its arrival at B the two clocks no longer synchronize, but the clock moved from A to B lags behind the other which has remained at B by $\frac{1}{2}Tv^2/c^2$ (up to magnitudes of fourth and higher order), T being the time occupied in the journey from A to B. It is at once apparent that this result still holds good if the clock moves from A to B in any polygonal line, and also when the points A and B coincide. If we assume that the result proved for a polygonal line is also valid for a continuously curved line, we arrive at this result: If one of two synchronous clocks at A is moved in a closed curve with constant velocity until it returns to A, the journey lasting T seconds, then by the clock which has remained at rest the travelled clock on its arrival at A will be $\frac{1}{2}Tv^2/c^2$ seconds slow".

Now we have indeed situations quite different from S.1, because only two clocks are involved. In *Situation 2* (S.2), A and B are initially separated and synchronized by an exchange of light signals. They are brought together by, as EINSTEIN expresses it, A moving and B remaining at rest. *Situation 3* (S.3) is not the same, in it A and B initially coincide and so may be directly synchronized. B then remains at rest, while A moves off to return eventually to its initial position.

2. SOME OBJECTIONS

In S.2 and S.3, what is there to justify EINSTEIN's designation of one clock as the travelled clock and the other as the clock which remains at rest? If we are to believe that all motion is relative, we may equally well say that A has remained at rest while B has moved. Therefore, say some, we could deduce by the same logic that B should lag behind A. This is the clock paradox. The description seems to be reserved by common usage for S.3 but clearly S.2 is not less paradoxical in its outcome. EINSTEIN offered no comment on this point in his 1905 paper, which is perhaps surprising, especially as the fact that the special theory of relativity applies only when the frame of reference is an inertial frame is nowhere emphasized in that paper.

The result formed the basis of some discussion, in the early days of the theory, of its internal consistency. (LANGEVIN 1911, VON LAUE 1912, LORENTZ 1914, EINSTEIN 1918). More recently, DINGLE (1956a, b) and CULLWICK (1957) have concluded that EINSTEIN's prediction of the outcome of S.3 is incorrect, and that the correct solution is that A and B will be found to read the same. CULLWICK concludes that even in S.1, C and B will be found to read the same, since the orthodox solution $t = T$ and $t' = T/\beta$ implies a lack of reciprocity between K and K'. We have already disposed of this objection in Section 1. DINGLE accepts EINSTEIN's solution for S.1, but not for S.3, and bases his argument on the following syllogism (DINGLE 1957b):

According to the postulate of relativity, if two bodies (for example, two identical clocks) separate and reunite, there is no observable phenomenon that will show in an absolute sense that one rather than the other has moved. If on reunion one clock were retarded by an amount depending on their relative motion, and the other not, that phenomenon would show that the first had moved, and not the second. Hence, if the postulate of relativity is true, the clocks must be retarded equally or not at all: in either case their readings will agree on reunion if they agreed at separation.

This expresses clearly the difficulty which many physicists must have felt about accepting EINSTEIN's result. DINGLE has skilfully defended this point of view in a series of articles (1956a, b; 1957a, b, c, d, e).

There can however be no doubt that DINGLE is wrong. That his argument is mistaken and his solution of S.3 incorrect (it is that both clocks will read T/β) has been pointed out by numerous critics (MCCREA 1956a, b; CRAWFORD 1957a, b; COCHRAN 1957b; FREMLIN 1957; BUILDER 1957a, b; DARWIN 1957). Support for EINSTEIN's solution had of course been forthcoming before the recent controversy (BORN 1922; EDDINGTON 1920, 1929; TOLMAN 1934; MCCREA 1951; IVES 1951; MOLLER 1952).

3. RESOLUTION OF THE PARADOX

If a short answer is to be given to the objections, it is that there is not equivalence between A and B in either S.2 or S.3, because A is accelerated and B is not. We shall

consider only S.2 in some detail because it can be discussed quite fully using simple mathematics, can easily be extended to apply to S.3, and shows most clearly the incorrectness of DINGLE's position.

Let us begin by specifying S.2 more precisely. A and B are separated by a distance X and are at rest in an inertial frame; that is, they are a relative rest in a region so remote from massive bodies that permanent gravitational fields are negligible and test particles move relative to the two clocks or to one another in straight lines with uniform velocity. The two clocks are synchronized by an exchange of light signals. Subsequently A, but not B, is accelerated out of rest in this inertial frame, reaching a velocity v in a time t_a which we may suppose to be negligible compared with T or T/β. That is, we suppose the acceleration to have been completed before the hand of clock A has moved appreciably from zero.

Since B remains in an inertial frame, we deduce at once, as in Section 1, that on meeting, B will record $T(=X/v)$ while A records T/β, a difference of approximately $\frac{1}{2}Tv^2/c^2$, with B ahead. According to DINGLE's interpretation of the principle of relativity, however, the motions of A and B are entirely equivalent throughout and A can be taken to be at rest, the same argument applying. This leads of course to the conclusion that, on meeting, A is ahead of B. DINGLE's theory therefore does lead to a paradoxical outcome for S.2, in that it gives two different answers to the question of what the clocks will actually *read* when they meet. DINGLE has published nothing on this problem, but the writer understands from a private communication that he hopes to deal with S.2 later.

Of course it is perfectly correct to insist that it must be possible to predict the readings of the two clocks by a calculation based on a reference frame in which A is at rest and B moves. In this calculation however we cannot rely exclusively on the equations of the special theory, because this frame is not an inertial frame throughout. A rigorous calculation (an adaptation of that of MOLLER (1952)) is tedious and difficult to follow; the following approximate treatment gives the correct result as far as terms in v^2/c^2. It has been shown by EINSTEIN (1911), using quite simple arguments, that two clocks which are subject to a difference of gravitational potential q will run at rates which differ by a factor $(1+q/c^2)$, (the general theory of relativity (EINSTEIN 1916), gives the factor more precisely as $(1+2q/c^2)^{\frac{1}{2}}$), the clock in the region of higher potential running relatively fast. Now in S.2, for the time t_a during which A is accelerated, the equivalent of a gravitational field $g \approx v/t_a$ is experienced, since a fundamental postulate of the general theory is the equivalence of an acceleration and a gravitational field. Thus in the frame in which A is at rest (and B is falling freely), A and B differ in gravitational potential by an amount $q \approx gX$. Relative to A therefore, B runs at a rate $1+gX/c^2$ for a time t_a, so that at the end of the acceleration their readings are t_a (for A) and $t_a(1+gX/c^2)$ (for B), a difference of $t_agX/c^2 = Tv^2/c^2$ since $X = vT$. Note that this difference remains, with B ahead, even when $t_a \to 0$, for t_ag remains finite. After the acceleration, both clocks are in inertial frames, and, keeping our origin on A, B, now runs slow relative to A to the extent $\frac{1}{2}Tv^2/c^2$, so that as a net result, when the clocks meet B is *ahead* of A by an amount $+Tv^2/c^2 - \frac{1}{2}Tv^2/c^2$, in agreement with the result of the calculation based on the frame in which B remains at rest. This calculation is admittedly approximate, but the general method of approach is correct.

The extension of this argument to a situation similar to S.3 is fairly obvious. Let the time recorded by B between successive coincidences be $2T$; in this interval A

moves away a distance $X = vT$ in a straight line and returns. A calculation based on the inertial frame in which B remains at rest shows at once that at the second coincidence B is ahead by approximately $\frac{1}{2}(2T)v^2/c^2 = Tv^2/c^2$. The calculation based on the frame in which A remains at rest proceeds as follows. The initial acceleration occupies a time t_a which is negligible. B also records negligible time, for although the gravitational field $g \approx v/t_a$ is large, the difference of gravitational potential is negligible because A and B are at this stage separated by a negligible distance. Keeping the origin on A, B now recedes and runs slow, so that when the reversal of the motion is about to begin, B is slow by approximately $\frac{1}{2}Tv^2/c^2$. The reversal of the motion occupies a time $2t_a$ for A, which is negligible, but the difference of gravitational potential makes the interval $2t_a gX/c^2 = 2Tv^2/c^2$ for B, so that B is now ahead by $(2T - \frac{1}{2}T)v^2/c^2$. During the return journey B runs slow by $\frac{1}{2}Tv^2/c^2$, and as a net result is ahead by Tv^2/c^2 in the end, the final deceleration being unimportant for the same reason as the initial acceleration was unimportant.

There remains however a very real difficulty of understanding. How can it come about that a gravitational field, which for B is non-existent, can cause B to run ahead by a finite amount and accomplish this in a time which according to A is negligible? Or to put the question in another way, is the eventual difference of clock reading in S.2 "caused" by the relative velocity (B's explanation), or is it "caused" by a combination of the initially-acting gravitational field and the relative velocity (A's explanation)? In an attempt to answer this question (or rather to show that there is no answer to such a question) let us look again at S.1, which falls entirely within the confines of the special theory and on which DINGLE and his critics agree. Does C really run slow in going from A to B (the explanation of A and B), or is the distance between A and B really X/β, and A and B really not correctly synchronized (C's explanation)? A quotation from a paper by EINSTEIN (1916) should serve to drive home this point: "All our space–time verifications invariably amount to a determination of space–time coincidences. . . . Moreover, the results of our measurements are nothing but verifications of such meetings of the material points of our measuring instruments with other material points, coincidences between the hands of a clock and points on the clock dial, and observed point-events happening at the same place at the same time. The introduction of a system of reference serves no other purpose than to facilitate the description of the totality of such coincidences". Space and time have no physical objectivity in relativity theory. While different observers must agree as to what two clocks actually read at a coincidence, their explanations (in terms of space, time and motion) of how the clocks came to have these readings may be completely different without one explanation being more, or less, correct than the other. This fact may well itself merit the adjective "paradoxical", but there is no other possible answer to either of the above two questions. The theory of relativity was a revolution in philosophy as well as in physics.[1]

4. The Experimental Evidence

The question whether two clocks, which are synchronized when they are together, will necessarily agree if they are separated and subsequently brought together again

[1] Those who agree in disagreeing with DINGLE would not all accept the philosophical position summarised in the above paragraph. In an article in *Austral. J. Phys.*, **11**, Sept. 1958, BUILDER re-introduces the ether hypothesis as a causal explanation of the phenomena we are considering.

is of course one that can in principle at least be answered by direct experiment. At the present time a charged particle, or a group of such particles, provide the only clock which can move in the laboratory with a speed comparable with the speed of light. It is of course well known that μ-mesons, whose mean life when at rest in the laboratory is $\tau = (2{\cdot}15 \pm 0{\cdot}07) \times 10^{-6}$ sec., can reach the Earth's surface from a height exceeding 10 km, despite the fact that their journey, in our frame of reference, requires about 3×10^{-5} sec. Our explanation is of course that the time elapsed in the mesons' rest system is less than this by a factor β, and β may be as great as 10 or more. This result is relevant to S.1, but not immediately relevant to S.2 and S.3. Measurements by DURBIN, LOAR and HAVENS (1952), using π-mesons, provide direct and fairly accurate experimental confirmation of the outcome of S.1. Charged π-mesons at rest in the laboratory have a mean life $\tau = (2{\cdot}56 \pm 0{\cdot}1) \times 10^{-8}$ sec. In the experiment a beam of π-mesons of energy $73{\cdot}4 \pm 2{\cdot}0$ MeV was projected into the laboratory, and its attenuation with distance measured. The arrangement of apparatus is shown schematically in Fig. 2. A meson beam is defined by three-fold coin-

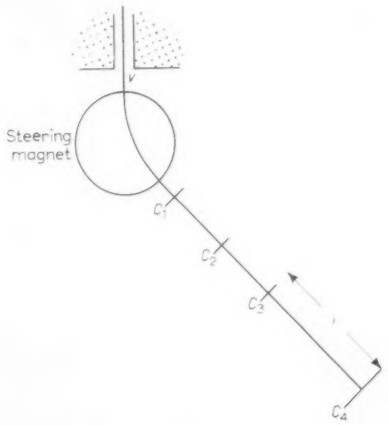


Fig. 2.

cidences in the scintillation counters C_1, C_2 and C_3. A fourth counter can be set at a variable distance l from C_3, and the fourfold coincidences measure the proportion of π-mesons surviving the distance l. It was confirmed that this proportion is given by $\exp[-l/(v\tau\beta)]$, so that the time recorded by the mesons in travelling from C_3 to C_4 is $l/(v\beta)$. In the laboratory system it is of course l/v.

That such an experiment could readily be modified to become the experimental equivalent of S.3 has been pointed out by ISAAK (1957), COCHRAN (1957a, b), GOUDSMIT (private communication) and no doubt by others. Really all that is required is a measurement of the proportion of mesons surviving flight on a curved path of given length, as compared with the proportion surviving flight with the same velocity on a straight line of the same length. Clearly the proportion surviving flight in a complete circle can be unambiguously deduced from the proportion surviving flight on any segment of the circle. There is no reason to doubt that the answer would be as found by DURBIN, LOAR and HAVENS. (See for example, JANOSSY, 1948).

The most striking and conclusive form of the experiment, although one possibly

difficult to achieve for technical reasons, would be as indicated in Fig. 3. The meson beam is deflected by a magnet to pass again through the fourth counter, which would be wide enough to intercept all returning particles. These would traverse a distance L of about 10 metres. Particles which did not return would be recorded on a cathode ray tube as a single pulse, those surviving the "space journey" would be recorded as a double pulse. The constant separation of these double pulses would give the time of flight in the laboratory, which should agree with L/v. The proportion of mesons surviving the journey would give the time of flight in the mesons' rest system. The counter C_4 and its associated electronic equipment is the clock B, the mesons are the clock A.

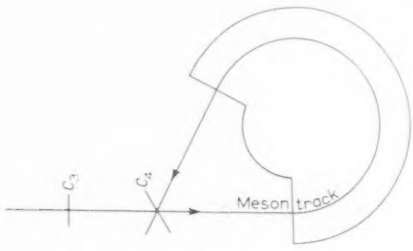


Fig. 3.

5. Space Travel in a Gravitational Field

It has been suggested by Singer (1957), and according to newspaper reports also by Russian physicists, that the prediction that "a moving clock runs slow" might be tested by carrying a clock in an Earth satellite. This is an experiment which is obviously nearer to achievement than more ambitious forms of space travel (for anything but mesons, etc.!), but the effect to be sought is extremely small. The experiment would be of considerable interest, as it would provide a test of the theory of relativity in the same sense as the much-looked-for gravitational red shift in the light from a star serves as a test. The rate of a clock relative to a clock resting on the Earth's surface (which we denote by $\mathrm{d}\tau'/\mathrm{d}\tau$) cannot be predicted from the special theory. Since both clocks are in the Earth's gravitational field the general theory should be used. In fact, depending on the radius of its orbit, a satellite clock may either lose *or gain* relative to an Earth clock. K. Schwarzschild's exterior solution (1916) for the motion of a small mass in a spherically symmetrical gravitational field gives, to a sufficient approximation,

$$\frac{\mathrm{d}\tau'}{\mathrm{d}\tau} = 1 + \frac{GM}{Rc^2} - \frac{GM}{rc^2} - \frac{1}{2}\frac{v^2}{c^2} \qquad \ldots\ldots (5.1)$$

where $G =$ gravitational constant, $M =$ mass of Earth, $R =$ radius of Earth, $r =$ distance of the mass from the Earth's centre and v is its velocity relative to the Earth's centre. For a satellite moving in a circular orbit of radius r we also have to a sufficient approximation that $v^2/r = GM/r^2$, so that

$$\frac{\mathrm{d}\tau'}{\mathrm{d}\tau} = 1 + \frac{GM}{Rc^2}\left(1 - \frac{3R}{2r}\right) \qquad \ldots\ldots (5.2)$$

This result can also be derived from McCREA's analysis (1956c) of a very similar situation. The acceleration of gravity at the Earth's surface is given by $g = GM/R^2$, so that the deviation per second, $\Delta = d\tau'/d\tau - 1$, of a satellite clock from an earth clock, is given by

$$\Delta = 2\Delta_0\left(1 - \frac{3R}{2r}\right) \quad \ldots . (5.3)$$

where

$$\Delta_0 = \tfrac{1}{2}gR/c^2 \quad \ldots . (5.4)$$

For an orbit of minimum radius $r \approx R$ and $\Delta = -\Delta_0$, representing a loss by the satellite clock of about one part in 3×10^9. For a satellite clock in an orbit of radius $r = 3R/2$ however, $\Delta = 0$; for any orbit of greater radius the satellite clock gains, the limiting value of Δ being $+2\Delta_0$ (SINGER, 1956).

For an orbit of radius $r_1 \simeq 6{\cdot}5R$, whose period is 24 hours, the value of Δ is $+1{\cdot}54\,\Delta_0$. This case is of particular interest, because a clock in an equatorial orbit of this period could be kept under continual observation from one place on the Earth, and the value of Δ determined experimentally by measuring the beat frequency between signals controlled by two atomic clocks. These can be constructed to measure time to one part in 10^{10}, (see, for example, LYONS, 1957) so the effect is not obviously beyond detection. There would be no correction for DOPPLER shift, for in this situation the two clocks are relatively at rest. It is therefore not at first obvious why their relative rate should not be determined simply by their difference of gravitational potential, which would give $\Delta = GM(1/R - 1/r_1)/c^2 = 2\Delta_0(1 - R/r_1)$. The additional term $-\Delta_0 R/r_1$ required by Equation 5.3 can however be ascribed to the fact that in this situation both clocks are at rest in a *rotating* frame of reference, (EINSTEIN, 1916). Once again, a different explanation leading to the same value of Δ would be given by an outside observer who did not share the Earth's rotation (cf. Section 3).

It is amusing to find, in view of the controversy on the ageing of space travellers, that in the simplest form of space travel, the traveller ages *most*! For a clock Z which is thrown up in a uniform gravitational field, moves in a straight line, and returns *under its own weight* will record a greater time than a resting clock Y with which it is synchronized at the point of release. This prediction of the general theory can also be understood using the special theory and the principle of equivalence. For the situation is entirely equivalent to the following, which involves no permanent gravitational field: B remains at rest in an inertial frame, A passes with a high velocity and is synchronized at the instant of coincidence. Thereafter A is subject to a constant acceleration towards B and eventually returns with reversed velocity. This situation sufficiently resembles S.3 for us to say at once that B will be ahead of A at the second coincidence. In a permanent gravitational field however the roles of "resting" and "travelling" clocks are interchanged relative to the above situation; Y $\equiv$ A and Z $\equiv$ B, for Z falling freely in the gravitational field is the equivalent of B at rest in an inertial frame, and Y at rest in the gravitational field is the equivallent of A undergoing a constant acceleration in an inertial frame. To predict the clock readings we may use a frame with origin on Y and the general theory, or a frame with origin on Z and the special theory.

A calculation based on Equation 5.1 (that is, with origin on Y) shows that a clock which is projected up with velocity U in the non-uniform field of the Earth, rises a

distance D from the centre of the Earth and returns under its own weight in time T, will show a reading greater than an Earth clock by a fractional amount

$$\Delta = 2\Delta_0(1-3R/D+4U/(gT))$$

For $D \gg R$, $\Delta = +2\Delta_0$. To avoid confusion it should be pointed out that if U is much greater than the escape velocity, that is $U^2 \gg GM/R$, the gravitational terms in Equation 5.1 become negligible and $d\tau'/d\tau \simeq 1-\frac{1}{2}v^2/c^2 < 1$, so that Δ is negative and the space traveller ages least. (In this situation of course rocket power would have to be used to bring the traveller back).

Let us imagine a rocket launching platform above the Earth's atmosphere, yet firmly attached to the Earth, and two space-travellers sent off simultaneously, one in an orbit round the earth and the other vertically upwards. The latter is given a velocity less than the escape velocity, just sufficient to allow him to return at the instant when the former has completed his first orbit. It would be found that the traveller in the satellite had aged less than a member of the launching crew, while the other traveller had aged more! This problem is considered in detail by MIKHAIL (1952). The differences would be too small to encourage or discourage either traveller, but I trust that this glimpse of strange happenings in the world of the future to some extent meets the Editor's suggestion that each article should conclude with a vista!

REFERENCES

BORN, M.	1922	*Einstein's Theory of Relativity* (Methuen, London).
BUILDER, G.	1957a	*Austral. J. Phys.*, **10**, 246.
	1957b	*Austral. J. Phys.*, **10**, 424.
COCHRAN, W.	1957a	*Nature (London)*, **179**, 977.
	1957b	*Proc. Cambridge Phil. Soc.*, **53**, 646.
CRAWFORD, F. S.	1957a	*Nature (London)* **179**, 35.
	1957b	*Nature (London)*, **179**, 1071.
CULLWICK, E. G.	1957	*Electromagnetism and Relativity* (Longmans, Green & Co., London).
DARWIN, C.	1957	*Nature (London)*, **180**, 976.
DINGLE, H.	1956a	*Nature (London)*, **177**, 782.
	1956b	*Nature (London)*, **178**, 680.
	1956c	*Proc. Phys. Soc. (London)*, A**69**, 925.
	1957a	*Nature (London)*, **179**, 1129.
	1957b	*Nature (London)*, **179**, 1242.
	1957c	*Nature (London)*, **180**, 500.
	1957d	*Nature (London)*, **180**, 1275.
	1957e	*Austral. J. Phys.*, **10**, 418.
DURBIN, R. P., LOAR, H. H. and HAVENS, W. W.	1952	*Phys. Rev.*, **88**, 179.
EDDINGTON, A. S.	1920	*Space, Time and Gravitation* (Cambridge University Press).
	1929	*The Nature of the Physical World* (Cambridge University Press).
EINSTEIN, A.	1905	*Ann. Phys. (Leipzig)*, **17**, 891.
	1911	*Ann. Phys. (Leipzig)*, **35**, 898.
	1916	*Ann. Phys. (Leipzig)*, **49**, 769.
	1918	*Naturwissenschaften*, **6**, 697.
FREMLIN, J. H.	1957	*Nature (London)*, **130**, 499.
ISAAK, G. R.	1957	*Austral. J. Phys.*, **10**, 207.
IVES, H. E.	1951	*Nature (London)*, **168**, 246.
JANOSSY, L.	1948	*Cosmic Rays* (Oxford University Press).
VON LAUE, M.	1912	*Phys. Zeitschrift*, **13**, 118.
LANGEVIN, P.	1911	*Scientia (Asso)*, **10**, 31.

LYONS, H.	1957	*Scientific American*, **196** (2), 71.
LORENTZ, H. A.	1914	*Das Relativitätsprinzip* (Leipzig).
MCCREA, W. H.	1951	*Nature (London)*, **167**, 680.
	1956a	*Nature (London)*, **177**, 784.
	1956b	*Nature (London)*, **178**, 681.
	1956c	*Helv. phys. Acta, Supp.* IV, p. 121.
MIKHAIL, F. I.	1952	*Proc. Cambridge Phil. Soc.*, **48**, 698.
MOLLER, C.	1952	*The Theory of Relativity* (Clarendon Press, Oxford).
SCHWARZSCHILD, K.	1916	*S.B. Akad. Wiss. Berlin*, p. 189.
SINGER, S. F.	1956	*Phys. Rev.* **104**, 11.
	1957	*Nature (London)*, **179**, 977.
TOLMAN, R. C.	1934	*Relativity, Thermodynamics and Cosmology* (Clarendon Press, Oxford).

On the Clock Paradox

V. Fock

Academician, Academy of Sciences, U.S.S.R.
Physical Institute, University of Leningrad, U.S.S.R.

The Clock Paradox is not in our view a particularly abstruse question, but it has been very widely discussed in the literature and some not wholly satisfactory resolutions have been given.

The paradox arises from using incorrectly the concept of relative motion and from ignoring the distinction between inertial and non-inertial frames of reference. It consists of the following[1]:

We imagine a clock A to be at rest in some inertial frame of reference. Let a clock B move past A with a constant velocity v and then, after it has covered a certain distance, let it be subjected to a negative acceleration which reverses its speed so that it again passes A, this time with velocity $-v$. At the two instants at which B is travelling past A the readings of the two clocks can be compared directly without the intermediary of light signals. Such a comparison will show that clock B is running slow compared to A, or at least this is the result of applying the expression (1) for proper time.

Now, motion is relative. Therefore clock B could also be thought of as at rest. Then the other clock, A, must be looked upon as first receding from B with uniform speed and then approaching again so that the same equations[2] appear to predict also that A should be slow compared to B, in contradiction with the previous result.

The difference in the readings of two clocks situated at the same point in space is an absolute and objective fact. It depends neither on the frame of reference nor on the method of description. Therefore, as long as they are correct, all methods of description must lead to the same result. The contradiction obtained shows that an error was made somewhere in the discussion.

It is not difficult to see that the error consists in failing to take into consideration the fact that in the imagined experiment the clocks A and B were subjected to different physical conditions. Clock A did not receive acceleration and experienced no impulses whereas clock B was accelerated; it experienced an impulse which reversed the direction of its velocity. In other words the error consisted in assuming that the two frames of references connected with clock A and clock B respectively were to be treated as equivalent, which they are not; only the frame of A is inertial.

This is the qualitative resolution of the paradox. A quantitative theory must be capable of calculating the reading of a clock in accelerated motion. It has been noted that the so-called proper time τ defined by the equation[3]

[1] A detailed treatment of the relativistic problems, with all the calculations involved, and from which also the following discussion has been taken, is contained in my recent book which was published in Russian in 1955 and will soon appear in English translation (V. Fock, *The Theory of Space Time and Gravitation*; translated by N. Kemmer, F.R.S.; Pergamon Press, London and New York, 1959).

[2] See Section 14, l.c. (i.e. Footnote 1).

[3] See Footnote 2.

$$\tau = \int_0^t \sqrt{\left(1 - \frac{v^2}{c^2}\right)}\, dt, \tag{1}$$

only has a simple physical significance in the case of a constant velocity v. Then it measures the duration in its "own" frame of reference, of a process localized at a point moving with this velocity v; in this case therefore equation (1) determines the reading of a moving clock. If, however, the velocity v is variable, such an interpretation of (1) is rather questionable and the impermissible application of this equation to accelerated motion is the immediate origin of the contradiction.

By what can equation (1) be replaced for accelerated motion? In the first place one should be quite clear that there is no equation that would give the reading of a clock for all possible accelerated motions. (This circumstance is hardly ever emphasized in discussions of the clock paradox.) It has previously been remarked[4] that no theory can predict how a clock will behave when subjected to impulses and arbitrary accelerations, unless the details of the clock's construction are taken into account.

It is, however, possible to introduce the hypothesis that in those cases where the acceleration is caused by a gravitational field the reading of a clock in free motion in that field will be given by the expression

$$\tau = \frac{1}{c} \int_0^t \sqrt{(g_{00} + 2g_{0i}\dot{x}_i + g_{ik}\dot{x}_i\dot{x}_k)}\, dt, \tag{2}$$

i.e. by the integral which has to be a maximum for free motion. An argument in favour of this hypothesis is that the gravitational field, and only the gravitational field, has the property of pervading all bodies and of acting on all parts in proportion to their masses.

If this hypothesis is accepted there is no longer any possibility of a paradox. For now the reading of clock A is obtained by evaluating the integral in (2) along the path of clock A just as the reading for B is given by the corresponding evaluation along its path. Both integrals so obtained are invariant with respect to any change of variables of integration, i.e. for any transformation of space and time coordinates. The use of different frames of reference, inertial or non-inertial, is equivalent to evaluating one and the same integral with different variables. This fact obviously excludes the possibility of any discrepancy in the results.

One should note that in resolving the clock paradox we have deliberately differed from the traditional discussion in not using the "Principle of Equivalence"; because of the approximate nature of this principle,[5] its use could leave a doubt as to whether the paradox was really resolved in all approximations.

We now calculate approximately the reading of the accelerated clock B on the basis of the above-mentioned special hypothesis. We use the Newtonian approximation to the expression for ds^2 and perform the calculation in the inertial frame in which clock A is at rest. We have then

$$\tau = \int_0^t \left\{1 - \frac{1}{c^2}\left(\tfrac{1}{2}r^2 + U\right)\right\} dt. \tag{3}$$

[4] At the end of Section 14, l.c. (Footnote 1).

[5] See section 61, l.c. (Footnote 1).

Let U_0 be the constant value of U at the position of A. If B passes A for the first time at $t = 0$ and then, when returning, again at $t = T$ the difference in the readings of A between the passages is

$$\tau_A = \int_0^T \left(1 - \frac{U_0}{c^2}\right) dt = \left(1 - \frac{U_0}{c}\right) T. \tag{4}$$

Here the influence of the gravitational potential on the rate of A is taken into account. For the difference of the two readings of B over the same length of time we get

$$\tau_B = \int_0^T \left\{1 - \frac{1}{c^2}(\tfrac{1}{2}v^2 + U)\right\} dt \quad . \tag{5}$$

Therefore B will be slow compared to A by the amount

$$\tau_A - \tau_B = \frac{1}{c^2} \int_0^T (\tfrac{1}{2}v^2 + U - U_0)\, dt, \tag{6}$$

where the integral must be taken along the trajectory of clock B. By the law of conservation of energy we have

$$\tfrac{1}{2}v^2 - U = \tfrac{1}{2}v_0^2 - U_0, \tag{7}$$

where v_0 is the value of the speed of B at the point where $U = U_0$. The relation (7) enables us to write (6) in various different ways, for example, as

$$\tau_A - \tau_B = \frac{1}{c^2} \int_0^T (\tfrac{1}{2}v_0^2 + 2U - 2U_0)\, dt. \tag{8}$$

We now assume for the potential U that in the region considered it has the form

$$\left.\begin{aligned} U &= U_0 && (\text{for } x < x_1), \\ U &= U_0 + g(x_1 - x) && (\text{for } x > x_1). \end{aligned}\right\} \tag{9}$$

Let clock A always be at the origin of coordinates and let B move along the x-axis. The x coordinate of B will then be

$$x = v_0 t \qquad (\text{for } t < t_1), \tag{10}$$

$$x = x_1 + v_0(t - t_1) - \tfrac{1}{2}g(t - t_1)^2 \qquad (\text{for } t_1 < t < t_2), \tag{11}$$

$$x = x - v_0(t - t_2) \qquad (\text{for } t < t_2). \tag{12}$$

Here t_1 and t_2 are the times at which B passes through the point $x = x_1$ on its motion to and fro. These times are

$$t_1 = \frac{x_1}{v_0} \text{ and } t_2 = t_1 + \frac{2v_0}{g} = t_1 + t^*, \tag{13}$$

where

$$t^* = \frac{2v_0}{g} \tag{14}$$

is the duration of the uniformly accelerated motion. The time of return of B to $x = 0$ is, by (12)

$$T = t_2 + \frac{x_1}{v_0} = t_2 + t_1. \tag{15}$$

In calculating the integral it is most convenient to use (8) because in it the quantity $\frac{1}{2}v^2$ is constant and the difference $2U - 2U_0$ differs from zero only in the region $t_1 < t < t_2$ where (11) is valid. An elementary calculation using (14) gives

$$\tau_A - \tau_B = \frac{{v_0}^2}{c^2}(\tfrac{1}{2}T - \tfrac{2}{3}t^*). \tag{16}$$

If we had unjustifiably used equation (1) we would have obtained (16) without the term in t^*. This term gives the correction for acceleration. It follows from (16) that if the duration of the accelerated motion is $\frac{3}{4}T$, the clock B is not observed to be slow at all, and for $t^* = T$ it is even fast. It should incidentally not be forgotten that (16) is not general, but only derived under rather special assumptions concerning the nature of the motion.

We have performed the calculation in the inertial frame of reference. To repeat it in a frame connected with clock B would have no sense for we should only be evaluating the same integrals in terms of different variables.

On the Design of Large Telescopes

G. M. Sisson

Sir Howard Grubb Parsons & Co. Ltd., Newcastle-upon-Tyne.

Summary

The article discusses some aspects of telescope engineering which are being thrown into prominence by modern trends. The difficulties due to seeing are briefly mentioned, after which follows a discussion of optical alignment problems including the guiding of a Schmidt Telescope. Some consideration relating to the support and thermal conditions of a large mirror are reviewed. Finally some details are given of a large Dual Purpose Telescope design, and the means by which the difficulties could be met are discussed including a brief description of a novel mirror support system employing an airbag for weight relief.

1. Atmospheric Turbulence

This article will not attempt to describe in detail the engineering problems of large telescopes, but will discuss in a general way some of the design aspects which are being thrown into prominence by modern trends and ideas. Naturally the more refined observing techniques now in use demand from the telescope higher perfection of image formation, and to achieve this more careful attention than was customary in the past is today given to the choice of an observatory site with good seeing conditions; indeed a minor branch of astronomical research has sprung up in connection with this single problem, (see Danjon 1955; Ellison, 1955; Rösch, 1955). While image imperfections due to high level atmospheric disturbances can only be mitigated by the choice of a good site, the observatory designer is left with the practical possibility of local air control in the region of the dome opening and within the dome and telescope tube.

The first telescope with a build-in air conditioning system as an original design feature was the St. Michel 75-inch reflector for the Observatoire de Haute Provence, in which the telescope incorporated a suction system drawing outside air down the tube in a manner designed to match the rest of the dome ventilating-arrangements; this has already been described (Sisson, 1957). Practical experience with a ventilation system drawing outside air through the slit in the dome and down a closed tube past the mirror has been gained at the Cambridge University Observatory where a comparatively simple arrangement was installed in the 36-inch reflector and dome. The improvement in image quality has been so marked that much may be expected from this technique in the future at observatories which are not normally blessed with good seeing.

A further refinement which will be introduced at the Pic du Midi Observatory in the future will be the conditioning of closed telescopes, such as refractors and Schmidt telescopes, by filling them with helium held at constant temperature and stirred by fans within the tube. The use of hydrogen was considered but thought to be too dangerous having regard to the valuable optical elements.

2. Optical Alignment

When the light reaching the telescope has been given the best possible conditions it is time to consider what will happen to it within the telescope itself. It may generally be assumed that the reflected light will enjoy the benefits of any air conditioning provided for the incident beam, and the value of this will be greatest over a distance of about one-third of the focal length of the primary after which it will matter less.

The telescope tube has to support a photographic plate or a secondary optical system of some sort, in a fixed relationship to the primary mirror at all useful attitudes of the instrument, and this it must do within certain mechanical limits. This is not a new problem, but an increasingly exacting criterion is now being imposed by modern observing techniques which make higher demands on a conventional optical arrangement. The same applies in the case of telescopes which use modern optical systems of high theoretical and practical performance, with their trend towards a shorter focal ratio and a wider field.

For example the Cambridge Schmidt Telescope has an aperture of 17 inches and a focal length of 64 inches providing a focal ratio of F3·8, and when this instrument was designed it was considered that a comparatively simple mechanical arrangement could be used to locate the photographic plate. This plate is 6 inches in diameter and the design target was to deform it spherical and hold it in the focal surface to within about a thousandth of an inch. However experience has now shown that a normal focusing plate will locate the best focus to within about half this tolerance and therefore a more exacting criterion could have been used with advantage. Provided the engineering is good, it is invariably the case that attempting closer tolerances involves increased cost and increased hazard, so that the designer must match the budget to the design, and the design to the required performance; unfortunately it is often difficult to stretch these links till they will join.

The guiding of a Schmidt Telescope is a question which has been considered by many and a number of ideas have been put into practice. Technically the ideal arrangement would be to guide at the prime focus in the centre of the field but few astronomers would agree to such a sacrifice. The next best thing is to guide just outside the edge of the photographic field still at the prime focus, this being the classical arrangement for parabolic reflectors, with the added advantage in the case of a Schmidt that the guide star is free from Seidel aberrations, and in particular from coma. The disadvantages of such a system are the asymmetrical obscuration introduced into the main beam, and the rather expensive and fussy arrangements necessary to bring the eye-piece system out to the edge of the tube while still allowing a measure of freedom of field search to pick up a guiding star. If a fixed guiding system is acceptable the difficulties are greatly reduced but the further disadvantage remains that the guider is several degrees off the centre of the field and field rotation may enter as a guiding error. Access to an eye-piece halfway up the telescope may well be inconvenient and there is no doubt that a guiding position near the lower end of the tube is the most comfortable arrangement. There are therefore compelling reasons for the adoption of a simple refracting telescope fixed to the main tube, and provided the relative flexures of all the mechanical parts are matched a simple arrangement can be relied on to give accurate guiding within about half a second of arc at all attitudes as has been found on the Cambridge Schmidt. (Fig. 1.).

A refinement of this system is to provide the possibility of adjusting the flexure of the guider relative to the main tube as has been done by ZEISS on the Hamburg Schmidt. In this case the objective end of the guider is carried upon cantilever rods whose effective length can be altered; it is not known what the maximum guiding error reaches on this telescope but practical results appear to be satisfactory. In view of the success of these simple arrangements it seems hardly worth while attempting to improve upon them for conventional Schmidt photography, but if an objective prism is being used there may be a case for considering alternative guiding systems.

A reflecting guider designed by Father JUNKES has been fitted by COX, HARGREAVES and THOMSON onto the 25–38 inch Schmidt Telescope built by them for the Vatican Observatory; this is an 8-inch reflecting telescope fixed to the main tube. A particular feature however is the use at the top end of the guider of two plane mirrors each in a

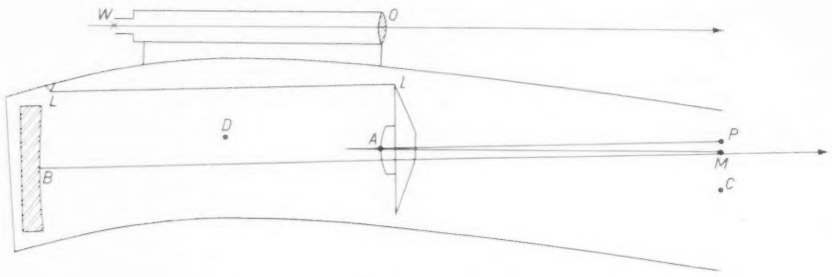


FIG. 1. Schmidt guiding with auxiliary refractor.

D	Declination axis	AM	Line of sight of schmidt
M	C of C of mirror	O	Guider objective
P	C of C of plateholder	W	Guider crosswires
C	C of correcting plate	OW	Line of sight of guider
A	C of photographic plate	LL	Focal length links
B	Pole of mirror	BM	is held parallel to AP by LL

Measured values for the Cambridge Schmidt:
CM = 0·016 inches MP = 0·008 inches

plane at 45° to the optical axis and parallel to each other. This pair of mirrors rotates as a single unit about an axis to scan a wide field of view in Declination, and one of them is pierced to pass the converging beam from the parabolic guiding mirror. This system allows the possibility of guiding at any angle up to 8° to the axis of the main telescope without inconvenient mechanical displacement of the guider telescope, a useful arrangement for objective prism work. It does however introduce a number of reflecting surfaces into the guiding system and the makers have had to give most careful consideration to the mechanical design to minimize the risk of tilt in any of the mirrors, due to structural flexure (Fig. 2).

For the Royal Greenwich Observatory at Herstmonceux a design study has recently been made on a special purpose Schmidt telescope to be equipped with a FEHRENBACH prism and work at F.10. Such an instrument would demand a highly accurate guiding system and a novel arrangement which seems to have many advantages. The telescope main mirror which is only 20 inches diameter and therefore very stiff, will have strapped onto its periphery a saddle in the form of a bracket carrying a 5-inch guiding mirror. The weight of mirror and bracket will of course be relieved by a counterweight and lever system to render it weightless so

far as the main mirror is concerned. The focal length of the guiding mirror will be twice that of the main mirror and a three mirror retroflector fixed to the spider which carries the main photographic plate, will return the guiding beam to a focus just behind the guiding mirror, which is pierced to pass this beam. It seems likely

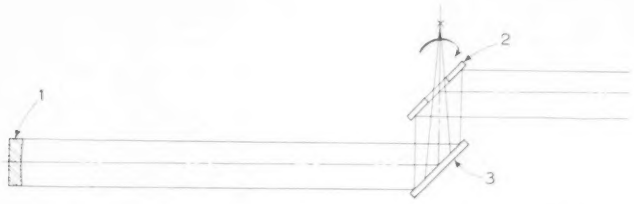


Fig. 2. A Schmidt Guider with sweep in Declination. Diagramatic arrangement.
1 = Guider Mirror. 2 & 3 = Plane reflectors which turn together.

that this will provide a very accurate and convenient guiding arrangement (Fig. 3). Finally, future experience may show that only photo-electric guiding methods can take full advantage of the Schmidt system applied to astrometry for which purpose it is so well suited.

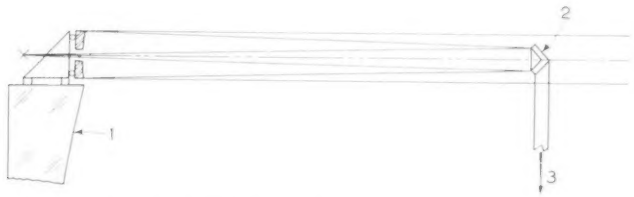


Fig. 3. A Schmidt Guider with retroflector.
1 = Main Schmidt mirror
2 = Retroflector.
3 = Spider support for Schmidt photographic plate.

3. The Mirror Cell

The figure of a large mirror may suffer deformations in the telescope due to changes in thermal conditions or unwanted changes in the forces exerted by its supports. A cellular mirror adapts itself more speedily to an ambient temperature change than a solid disc, but in doing so it suffers pattern distortion matching the ribbed structure of the back, while the solid mirror will show a general smooth distortion (Wellmann, 1954). If means for continuous general correction of shape are built into the support system, for example heating wires on the back of the disc as used by A. Couder, or a mechanical arrangement of some sort, then a very large solid disc is probably preferable to a ribbed one. These are questions which can really only be resolved by practical trial, but telescopes built to a price must be made to work when completed with a minimum of modification, and this imposes a severe limitation on the degree of novelty which can be introduced into their design. No other art is quite so hampered and it would be a good thing if means were found to do a measure of large-scale experimental work without committing a complete

telescope project. In any case the choice as between a ribbed or a solid disc is likely to be swayed by practical considerations such as availability in the size desired, and the significance of the weight of the mirror in the mounting design. A further influence will be the plan of action for working the mirror and supporting it during this process, in fact manufacturing and testing procedures play as important a rôle in governing decisions for the optical design as they do for the mechanical design.

At observatories where the night temperature follows a regular pattern of slow annual change from season to season, it is enough to provide a well insulated mirror cell in which the disc can be maintained by day close to its night working temperature; but in climates subject to large variations in the temper of the weather this arrangement is not good enough, although it is better than nothing and usually all that is provided. However, at Cambridge a temperature-conditioning system has been applied to the simple cell of the 36-inch reflector, by means of which cold air can be circulated within this cell when the telescope is out of use. A standard air refrigeration unit has been adapted to provide air at any desired temperature and this unit is connected by flexible pipes to the cell during the day; experience with the arrangement has been very satisfactory.

The support of a mirror disc within its cell is a difficult problem involving the application of controlled forces to the glass in both the axial and lateral directions. These two sets of forces are applied by mechanical systems which are sometimes combined and sometimes quite separate, and while it might be thought that experience would soon point the way to the simplest and best arrangement, this does not seem to have happened and many different opinions are heard. The fact that it is impossible to test in a comprehensive way any support system under laboratory conditions is one of the reasons for the diversity of opinion; for the only test of a mirror cell is performance in the telescope, a lengthy process the results of which are often hard to discover and sometimes hard to believe. It does seem certain that very few large mirrors have been carried upon their support system without obvious deformation when first tried out, but it also appears that experience in service has pointed the way to defects in each individual system so that each has eventually been brought up to an acceptable standard of performance. In this way the position has arisen in which many different arrangements have been made to work but none is acknowledged as ideal.

The logical aim of any system must to be apply to each part of the disc a force of the right amount in the right direction to render the particular part weightless, and therefore the method of using compound levers with counterweights to apply both axial and lateral forces in the plane of the centre of gravity of the mirror must be theoretically the nearest approach to an ideal yet practical condition. A ribbed disc lends itself well to such an arrangement and at Palomar the largest mirror in the world is satisfactorily carried upon a compound lever system exerting lateral forces in the plane of the centre of gravity, and axial forces at the back face of the ribs. But with a solid disc the position is quite different because the central plane of the mirror is not readily accessible and it is necessary to bore rather large holes into the back of the mirror, with consequent danger of fracture during boring, and some loss of thermal uniformity which may occasion pattern distortion in service due to temperature changes.

An example of this principle is the lateral support applied to the solid disc of the 2-metre Universal Schmidt telescope now being built by Zeiss at Jena for the new

Potsdam Observatory in which the levers apply their lateral force at the bottom of holes in the back of the mirror. In this case as at Palomar the axial force is applied to the back surface of the mirror and not at the bottom of the hole, which leads to some simplification of the lever support system and seems to be a sensible compromise. The holes in the back of this disc were successfully hollowed out by sand-blasting, a comparatively safe procedure if the disc is very well annealed and homogeneous. The annealing of a disc such as this with a thermal co-efficient of linear expansion of $4{\cdot}9 \times 10^{-6}$ per °C can be very good, as it certainly was in this case, but as the thermal co-efficient is reduced the difficulty of casting and annealing a large glass disc increases sharply, and a boring operation on this scale would often not be an acceptable risk for a disc with a co-efficient of less than 3×10^{-6} per °C. The lateral support must then be applied around the circumference of the mirror and axial support applied quite separately to the back face.

For large mirrors this axial support has always been provided by an appropriate number of pads arranged in rings below the back face of the mirror, each possessing a little lateral freedom in order to accommodate temperature changes and avoid the exertion of any lateral forces which would distort the mirror. Generally each pad forms part of a lever system carrying a weight which ensures that the force applied by the pad is dependent upon the attitude of the telescope, in such a way that it is always matched to the duty it has to perform in carrying it's share of the mirror load. The fulcra of all these levers are attached to some mechanical part of the mirror cell and the loading at these points of attachment changes with telescope attitude.

This change of loading flexes the cell structure so that the fulcra change position by small amounts relative to each other, and unfortunately the most likely flexure is a general astigmatic distortion, which is the most disturbing to the mirror. If the lever system is working correctly the pads, being in contact with the glass disc, will not change level relative to each other, so that when the cell structure flexes each pivot will have to turn a little to accommodate the movement of the fulcrum supporting point, and in practice a second joint between the lever and the pad will also have to turn. The starting friction or "stiction" as it is sometimes called, in the bearings at these two points of rotation will show up as an error in the force being applied by the pad to the back of the mirror. By using small diameter ball bearings and conventional mechanical arrangements it is possible to reduce this error force to around one part in 200. Any individual supporting pad acting in an imperfect manner and applying a force which differs from that of its neighbours by an amount of this magnitude, will not produce a significant local deformation in the mirror; but when a large number of pads act together towards an astigmatic deformation, as described above, there is trouble. To assess the significance of this condition it is convenient to consider a pattern of error forces acting on the disc at the points where the support pads are. The forces over a central belt act one way while those in the crescents on either side act the other, thus tending to bend the disc along a diameter. The maximum flexure under this loading can be calculated and it appears that for mirrors around 74 inches diameter and 11 inches thick a support system of 12 pads with a friction error of 1 in 100 will not produce an astigmatic distortion greater than 1/20 wavelength of light. However, a 98-inch mirror 18 inches thick supported on 24 pads required a friction error as small as 1 in 220.

To avoid this difficulty a recent United States design employs a number of oil

filled bellows to support the mirror, each connected to a much smaller bellows acted upon by the usual lever and weight. The differential action of the bellows secures an effective lever ratio without friction and thus reduces the final contribution of the mechanical lever. The design for the Potsdam 2-metre telescope employs crossed spring flexures for the lever pivots; this arrangement eliminates all significant friction at the expense of increased complexity and some difficulties due to the rotational stiffness of the flexure elements themselves so that great care must be taken when adjusting them; for example a typical system for a pad carrying a 300 lb load using high tensile steel flexure pivots would have an error force rising to 1 part in 200 for a pad movement of 0·01 inches away from its nominal correct position; however it does afford a practical solution.

The method by which lateral support is applied around the circumference of a large mirror still affords wide scope for argument. Whatever method is employed the system must allow a reasonably free axial motion so that the lateral system cannot significantly affect the figure of the mirror when the cell structure flexes. The 75-inch mirror 7 inches thick in the St. Michel Reflector is simply located, and carried, upon three lead faced pads arranged at 120° around its circumference. There is abundant workshop evidence to show that a 74-inch mirror 11 inches thick supported on edge on two pads carrying its full weight is gravely deformed, but then no telescope ever does work in the horizontal position, and only experience can reveal the success of this very simple arrangement in practice.

All other large mirrors are supported by a system which divides the function between weight relief and location. Several 74-inch mirrors appear to be carried satisfactorily by a very simple arrangement of levers all acting on a number of brackets fixed to a belt bound round the periphery of the disc. In this case all the levers help to carry the weight at all attitudes, those in the upper half lifting by tension in the belt. The American design already referred to above employs oil-filled bellows to press radially on the lower half of the circumference under the action of a lever and counterweight arrangement. Mechanically this is convenient, but it seems to have the disadvantage of inefficiency in its application of the forces, since a few bellows near the bottom of the mirror have to carry practically the full weight. The ideal distribution of force around the circumference of a solid disc stood on edge has been investigated theoretically, and it appears that the most satisfactory arrangement would be to apply radial forces sinusoidally distributed around the periphery, rising from zero at either end of a horizontal diameter to a maximum tensile force at the top, and a maximum compressive force at the bottom; (Schwesinger, 1954).

There does not appear to have been any attempt to arrange the lateral supports for a large mirror in this way, but the discovery of glue which will stick metal to glass in a reliable way does open up new possibilities. There seems to be no reason why metal lugs should not be stuck to the glass as attachment points for links to pull or push radially.

If a number of levers are arranged around the edge of the mirror, each having a fulcrum with its axis tangential to the disc, the upper ends of these levers may be linked to lugs on the edge of the mirror and the lower ends loaded with counterweights. These levers all lying parallel to the optical axis of the mirror, will produce a sinusoidal force system around the periphery of the disc, in the manner which seems to be best suited to minimize distortion of figure when the disc is not lying on its

back, and it was intended to use an arrangement like this to offer lateral weight relief for the 98-inch mirror in the Isaac Newton Telescope design.

4. A Dual Purpose Telescope

As a conclusion to this article it is perhaps appropriate to describe and discuss here very briefly one of the designs which was developed in some detail for the Isaac Newton Telescope. The basic requirements were to provide a dual purpose telescope with the following characteristics.

Main Mirror. Aperture 98 inches. Thickness 18 inches. Solid Pyrex with expansion $2{\cdot}6 \times 10^{-6}$ per °C.
Central hole 14 inches diameter.
Front surface spherical 588 inches radius.
Back surface flat.
To be made from an existing disc presented by the McGregor Trust of Michigan U.S.A.

Correcting Plate. Aperture 80 inches. Thickness approx. 1 inch. Stirred white plate. To be easily removable.

Prime Focus. Photographic plates 12 inches square to be deformed in the usual way. Also prime focus spectrography.

Gregorian Focus. Removal of correcting plate and prime focus star plate to give a Gregorian system corrected on axis for spherical aberration.

Coudé Focus. Similar to the above system but with an alternative Gregorian mirror.

For conventional parabolic reflectors with an F.5 primary and normal Cassegrain or Coudé arrangements a Modified English mounting carrying a simple lattice girder type of tube will meet the customary alignment requirements up to sizes of 80 inches or so, beyond this it seems wise to use a mounting giving symmetrical support to the main tube of the telescope. Such a mounting allows the possibility of a balanced flexure or Serrurier tube construction first used with brilliant success for the 200-inch at Mt. Palomar.

It seems likely that in the future large telescopes will be constructed which take advantage of up-to-date servo techniques, and ensure continuous optical alignment by using subsidiary optical systems to render the important parts immune from mechanical flexure; it was intended to use such an arrangement for this design; (Sisson, 1954). The alignment of a secondary mirror or a photographic plate implies locating it relative to the optical axis of the main mirror within certain limits which may be referred to in terms of the mechanical flexures, Tilt, Axial Displacement and Lateral Displacement. In the case of this telescope it appeared that a common central supporting tube should be used to carry the plateholder at one end, and the Gregorian mirror at the other end. The criterion for lateral error, in the absence of tilt, for this support was the coma which would be introduced into the image by the highly aspherical Gregorian; and a limit of 0·025 inch lateral displacement was chosen to keep the image enlargement due to this cause below 1″ arc. Axial error was governed by the depth of focus with the instrument in use as a Schmidt Camera, and it was thought that the focal distance should be maintained within $\pm 0{\cdot}002$ inch; the proposed arrangement for achieving this automatically reduced the tilt to an insignificant amount.

It hardly seems possible to design a purely mechanical structure which will behave consistently within these small limits neither does it seem wise to attempt to do so, and therefore it was arranged that the common central support for the secondaries would be carried by a system of levers and counterweights so that it floated weightlessly and freely within the main tube. Axial location and tilt would be controlled by four fine screws driven by electric motors, three would be theoretically correct but four appears more practical; lateral location in two directions would be controlled by fine motions under servo motor control. Reference to Fig. 4 will show the proposals as sketched out in a practical way for incorporation in the telescope tube. The axial motors are energized through on–off microswitches actuated by compound rods having a thermal co-efficient of expansion matching that of the glass disc and lagged to give them a comparable time constant; these rods extend from the region of the spider down to the main mirror cell. The lateral servo motors are operated by a subsidiary auto-alignment optical system comprising a light source at one side of the spider, reflected in a small concave mirror fixed within the central hole of the main mirror, and finally brought to a focus again on the opposite side of the spider. This arrangement ensures that the common support for the secondaries and photographic plate is at all times carried in an attitude related directly to the front surface of the main mirror independently of any flexure in the tube structure itself. To maintain adequate alignment of the correcting plate a balanced flexure main tube design will prove fully adequate.

The disc for the Isaac Newton Telescope mirror is rather thicker than usual for its diameter. This is thermally acceptable in a disc with such a low expansion co-efficient, about $2{\cdot}6 \times 10^{-6}$ per °C, and it has the advantage of making the axial support problem less difficult to solve. In fact a simple lever and counterweight arrangement employing 24 pads has been worked out as shown in the table.

Table 1. *Support system for a mirror* 98 *inches diameter F*.3, *having a thickness at the edge of* 16 *inches and a central hole of* 22 *inches diameter. All pads have* 9 *inches diameter.*

Ring position	Radius in inches	No. of floating pads	No. of fixed pads	Load per pad (lb)	Load per ring (lb)
Inner	18	4	Nil	442	1770
Middle	31	8	Nil	365	2920
Outer	44	9	3	433	3910
Total	—	21	3	—	8600

It was however intended to design the mirror cell in such a way that a new and untried support system would be installed in the first place, and only in the event of the failure of this new proposal would the conventional pad system by used. The new proposal relates to the axial support only and employs an old idea first put into practice by FOUCAULT who used a pig's bladder to carry a mirror.

The modern version of this seems to be a neoprene bag beneath the mirror with an automatic pressure adjustment according to the attitude of the mirror cell, and this arrangement is not difficult to realize in practice. When the 98-inch mirror was being figured it was at one stage tested on its back lying directly upon a circular rubber

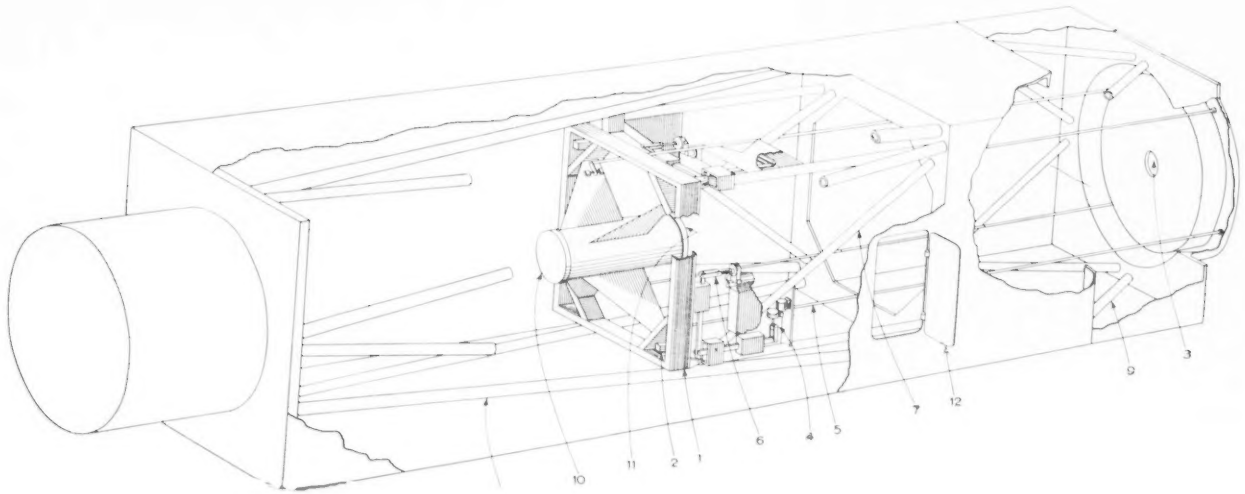


Fig. 4. Diagramatic drawing of dual purpose telescope tube.

1 = Spider frame. Balanced and floating carries plate holder and Gregorian.
2 = Optical Servo Unit. Aligns spider frame, i.e. Plate holder and Gregorian, with main mirror.
3 = Central mirror in main mirror used in conjunction with 2.
4 = Centring screw motions. They work in pairs controlling spider frame.
5 = Focusing and squaring on rods. Low expansion. Between back face of main mirror and corners of spider frame.
6 = Focusing screw motions controlling spider frame. Each works under control of nearest focusing rod 5.
7 = Triangular frame supporting spider frame through balance linkage.
8 = Main triangular frame supporting tube walls and correcting plate.
9 = Triangular frame supporting mirror cell. 8 & 9 sag equally.
10 = Gregorian.
11 = Plate holder. Removed for spectrographic work.
12 = Access door 5′ 0″ × 2′ 0″ to interior of tube for plate loading with tube horizontal. Also used for servicing, etc. Cat walk in tube.

bag; this was then filled with air at a pressure regulated by means of an escape tube lowered to a fixed depth beneath the surface of water standing in a tank. The rubber bag had to be centred fairly accurately to persuade the mirror to float upon it without tilt, but this was not found to be a critical factor or difficult to achieve. During the test the pressure in the bag was held a little above that required to carry the weight of the mirror, which was positively located by three fixed squaring on stops touching the front surface close to the edge of the disc. In making these arrangements it was found that after the mirror had been lifted by increasing pressure till it just touched the stops, it was possible to lower the escape tube a further $\frac{1}{4}$ inch before visible deformation of the image occurred (due to undesirable local pressure at the three stops). In other words the experiment showed that for this disc an axial error force of some 25 lb. can be tolerated at three local points near the circumference before the figure of the mirror is significantly deformed. If now a bag is placed beneath the mirror with three gaps bitten out of it near its circumference, it may be used to relieve the weight of the glass above at all points except where the three gaps appear, and in these places fixed pads may be arranged to carry their appropriate proportion of the weight, and locate the mirror. The area of the gaps may be arranged so that with the mirror horizontal each pad must carry (in round figures) a load of 200 lb, and the bag a distributed load of 8,000 lb to support the mirror between them, but it is important that the load on each pad should not differ from the nominal figure by more than 25 lb, if distortion of the mirror is to be avoided. In practice it has been assumed that the observed limit of 25 lb should be reduced to a target figure of 10 lb to be well on the safe side. The pressure inside the bag must then be always so accurate that it exerts the required lift within ± 30 lb corresponding to a pressure accuracy within 0·1 inch of water. To generate a pressure standard of the required amount a pivoted lever carrying a weight may be mounted inside the mirror cell, carried at the fulcrum by crossed springs and supported by a metal bellows connected by piping to the bag which supports the mirror. A number of ideas for a static system of this kind were investigated but in all cases the changes in pressure and temperature led to embarrassingly large changes of volume which were difficult to accommodate without prejudicing the accuracy of the pressure control. The necessity to seal the system permanently, or make good any leakage periodically, is also a disadvantage.

For all these reasons a dynamic system was chosen using a continuous supply of air brought into the mirror cell at a pressure of about 10 lb per square inch, either by means of a compressor mounted on the telescope tube or a flexible pipe led onto the tube alongside the electric cable run. The pressure controller is illustrated diagrammatically in Fig. 5. It will be seen that if the pressure rises and lifts the lever, the escape jet passes more air thus reducing the bag pressure till the force due to the bellows balances the load due to the weighted lever, and this load changes with attitude in the desired manner provided its maximum value has been correctly adjusted.

Such a device has been made and tested quite rigorously under realistic conditions in the laboratory with the following results.

Stability: The system was stable and did not hunt. The final parameters will be different from those under test, but no difficulty is anticipated in this respect.

Speed of response: This is adequate to meet the quick motion requirements on a large telescope for both rising and falling pressure.

Consistency: Large disturbances were found to be corrected consistently to within 0·020 inch of water of the nominal correct reading. Variation of $\pm 20\%$ in the input pressure resulted in changes of $\pm 0{\cdot}02$ inch in the controlled pressure.

Linearity: Over a pressure range of 15 to 30 inches of water the maximum observed departure from an ideal linear relation between applied weight and resulting pressure, was found to be $\pm 0{\cdot}024$ inch of water.

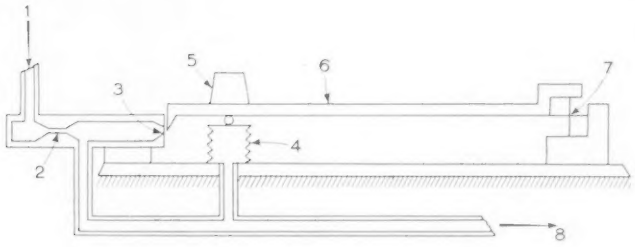


Fig. 5. Air Pressure Controller.

1 = Primary air supply.
2 = Primary control jet.
3 = Escape jet.
4 = Bellows.
5 = Weight on lever.
6 = Lever.
7 = Crossed spring fulcrum.
8 = Controlled pressure to airbag.

The errors listed above are not very much larger than the probable errors of observation so that the experimental work cannot be claimed to show how good the controller really is; but these and other test results show that a pressure controller of this type can fulfil the requirements for an air bag support beneath the 98-inch mirror.

Clearly the use of such a bag directly in contact with the glass would not be an acceptable arrangement because without some spacing the back of the mirror would be thermally insulated. The proposed design is shown in a sketch (Fig. 6), where it will be seen that a number of rubber balls of 2 inch diameter are used directly below the glass. The disposition of the balls will have polar symmetry and the spacing between them will be about 6 inches, the accurate positioning of the balls is not important. To integrate the load appropriate to each of them a thin metal sheet is needed between the balls and the bag. The fixed pads are pivoted on levers held fast against squaring on stops when the air supply is on, and the support in use; but when the air supply is shut off the mirror can settle down on the empty bag.

In connection with the criterion for pressure control it is worth mentioning that according to calculation an error in the air pressure within the bag, which results in a total lifting error of 51 lb, that is 17 lb on each of the locating pads, is enough to cause a flexure of 1×10^{-6} inch at the centre of the mirror. The first-order form of such a distortion will be spherical, and the experimental evidence suggests that only when this flexure reaches $1{\cdot}5 \times 10^{-6}$ inch does the test image start to show a triangular distortion with this particular mirror. There are a number of features of the proposal which cannot be dealt with here, but enough has been said to show that

the system holds promise for the future, in particular an arrangement of two or three concentric airbags would at once offer the possibility of trimming the figure of the mirror during use.

The above description shows how the immediate optical difficulties inherent in the Dual Telescope design might be met, and it remains to discuss the mounting for the complete telescope. In any large telescope design the Coudé system is nearly always found to be an embarrassment for a number of reasons. In the first place the 45° flat is a central stop in the converging beam from the primary, and if its obscuration is to be held reasonably small this places a limitation on the position of the Declination Axis along the main tube; this is particularly troublesome when the focal ratio of the main mirror is small as in this case. Then again the long optical

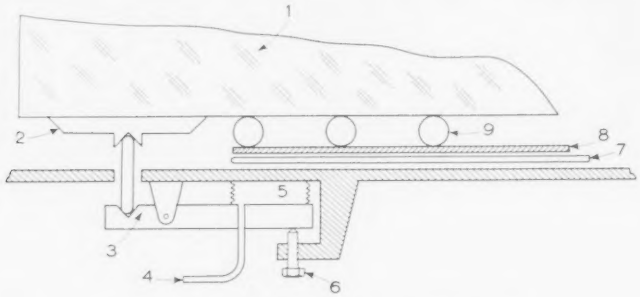


Fig. 6. Airbag support for 98-inch mirror.

1. Main mirror.
2. Locating pad.
3. Pivoted lever.
4. Air from primary supply.
5. Bellows acting on lever.
6. Squaring on stops.
7. Airbag.
8. Thin metal sheet.
9. Rubber balls.

path imposed by a good rigid mounting with a long Polar Axis, involves large secondaries if an acceptable Coudé focal ratio is to be maintained, and this tends to increase the obscuration. Finally if a conventional fork or yoke mounting is chosen then auxiliary mirror systems must be used near the Pole, and in all positions the slit in the tube is a mechanical weakness.

All these thoughts lead towards an offset fork as the most suitable mounting in spite of its very large size and weight (Fig. 7). Such a design using externally pressurized oil bearings for the Polar Axis seems to offer a practical solution to all these difficulties, and two alternative arrangements of Coudé focus were investigated. One of these was a conventional system bringing the focus down the Polar Axis in the usual way, but the other suggested by Dr. R. d'E. Atkinson, explored the possibilities of a narrow Coudé spectrograph room inside the dome observing space (Fig. 8). It will be seen from the figure that the Coudé focus is brought northwards up the polar axis centre line, and by this means a very favourable focal ratio of F.18 can be obtained, whereas a conventional Coudé focus at the lower end of the Polar Axis could not be arranged with a better ratio than F.45.

It would be tedious to attempt a discussion of the many practical difficulties foreseen in drawing up the proposals which have been rather inadequately described here.

The construction of this telescope would have broken much new ground on a

Spectrograph room
section thro' *XX*

Dewcap	80" long
Dome internal dia.	100 ft
Polar axis upper bearing	40 ft dia
Spectrograph chamber	6 ft wide internally 9" walls
Focal ratio	F. 20
Gregorian	20·6" aperture
Prime focus at slit	350" optical distance
First coudé flat	13" circular aperture
Collimator aperture	12"
Cameras	25" focus with Schmidt plate 50" " without " " 125" " " " "

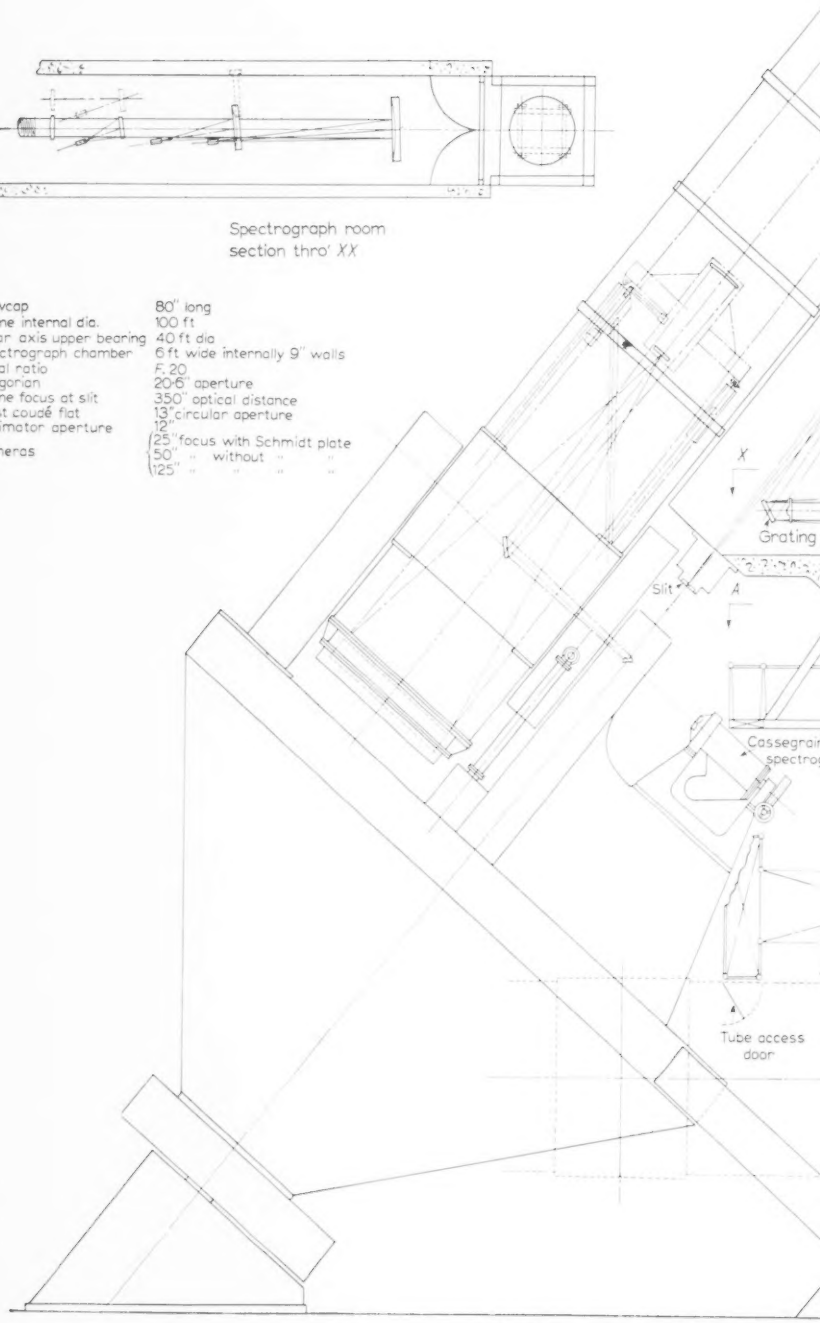
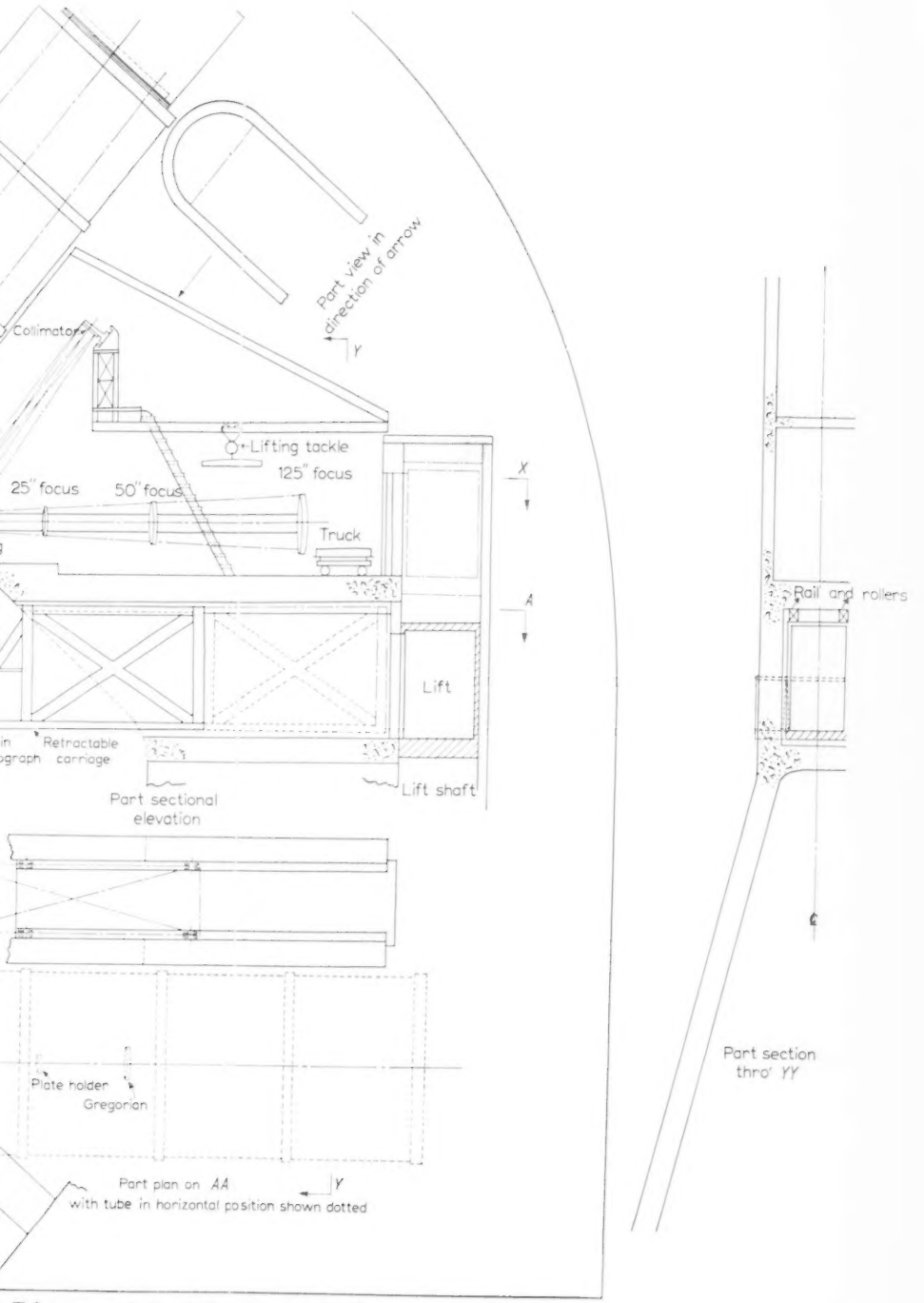


FIG. 8. Cross-section of Dual Purpose



e Telescope with Coudé focus to the North.

[facing page 104

VOL
3
196

Fig. 7. Model of Offset Fork Mounting.

grand scale, but it involved a number of uncertainties which were only capable of definite resolution by experience. For these and other reasons the Dual Purpose Isaac Newton Telescope design will not be used.

REFERENCES

Danjon, A.	1955	*Vistas in Astronomy*, ed. A. Beer, Pergamon Press, London; vol. I, p. 377.
Ellison, M. A.	1955	*Proc. Symposium on Astron. Optics, Manchester Univ.*, p. 293.
Rösch, J.	1955	*Proc. Symposium on Astron. Optics, Manchester Univ.*, p. 310.
Schwesinger, G.	1954	*J. Opt. Soc. Amer.*, **44**, 417.
Sisson, G. M.	1954	*Occ. Notes Roy. Astron. Soc.*, vol. 3, No. 17.
	1957	*Nature*, **179**, 937.
Wellmann, P.	1954	*Z. Astrophys.*, **35**, 104.

VOL
3
196

The Electronic Computer as an Astronomical Instrument

MARSHAL H. WRUBEL

Indiana University, Bloomington, Indiana, U.S.A.

SUMMARY

The enormous potential of the electronic computer as an instrument of observational and theoretical research is emphasized. Astronomical examples are used to illustrate the concept of programming and the flexibility of electronic computers. Specific recommendations are made for the effective use of these devices in astronomy.

1. INTRODUCTION

THERE are very few astronomers fortunate enough to work with paper and pencil alone. For most of us, there is a long path to be followed, from observations to conclusions, or from theory to prediction, before our work can be useful to ourselves and others. In some cases the data obtained in one night of observations may take weeks to analyse. This process may be so time-consuming that frequently only a fraction of the data are reduced before the astronomer is diverted by new problems. As a result much useful information gathers dust and never appears in the literature.

What can be more detrimental to astronomical enthusiasm than routine, dull, laborious computation! It has strangled some large programmes, it has delayed others and it frequently drives promising graduate students completely away from astronomy. Unless a large force of desk calculator operators is available (and this is unlikely when astronomers must hire assistants in competition with industry), the astronomer must do the work himself and spend valuable hours at menial tasks, instead of in creative thought.

With this obstacle in the path, one would think that astronomers would leap at the prospect of using high-speed computers. This has not been the case. (Astronomers are innately as conservative a group of scientists as one can find.) With the exception of a few widely dissimilar fields the use of computers has been timid and sporadic.

By and large, the principal reason has been inertia. Astronomers are vaguely aware of what electronic computers can do to help them, but few have taken the trouble to investigate further. For some reason, electronic machines are regarded either as mysterious or difficult to learn to use. In reality, with a little guidance, anyone familiar with desk calculators can be producing useful results with a high-speed machine in a matter of six weeks or less. While this may appear to be a long time to devote to learning a new technique, it is an investment which saves far more time in the end.

Naturally, astronomers cannot be expected to change their methods overnight. It will be the young astronomer, accustomed to learning new things, who will accept electronic computation in his stride. It is to these men, especially, that this article is addressed.

2. Simple Examples of Computer Techniques

Before discussing specific astronomical applications, let us examine the electronic computer in more detail. It has five principal parts: the input, into which information if fed; the output, out of which results eventually come; the memory, where information is retained for use during the computation; the arithmetic unit, in which the calculation is performed; and the control unit, which supervises the entire operation.

Each machine has a basic vocabulary of instructions which it is able to perform. These are the "words" of the "machine language". By assembling these instructions in the proper sequence, a series of operations can be performed on the data to solve a specific problem. This sequence is called the "programme" and the process of constructing it is called "programming". The instructions of the programme must be written in a (usually numerical) code that the machine can understand. This process is called "coding".

Once the correct sequence has been constructed for a particular problem, it can be used again and again on different data. For example, once a programme for reducing photoelectric observations has been written and coded for a particular type of computer, it can be used without further programming, to reduce as much data as desired. It can also be used on any similar machine anywhere in the world.

In practice, the entire programme is stored in the memory. Once the execution of the programme is begun it continues automatically. After the calculation has been completed for one set of data, the results are punched or printed, and new data are automatically fed in.

The programme *library* is one of the most important features of any computer installation. Programmes of general interest have been developed for all the standard machines and many of them have astronomical applications. For example, least-squares curve fitting programmes are available, and to use them it is only necessary to punch the data on cards or tape in the proper form. Standard procedures of interpolation, quadrature, and many statistical operations are already available and can be used by astronomers with very little additional work.

Purely astronomical applications, however, will have to be programmed by the astronomers themselves. If they wait for someone else to do it, it will never be done.

Fortunately, many techniques are being developed to make programming as simple as possible. The tendency is toward making the machine do most of the work. While the professional programmer regards these short-cuts with some disdain, the beginner would do well to learn them. Although programmes written by these methods may not make the most efficient use of the machine, they lead to workable error-free programmes in a much shorter time.

One such approach is called "automatic programming" and a typical example is FORTRAN, developed by IBM for its Type 704 computer. FORTRAN is very attractive to the scientists because programming is done in a language very similar to algebra. For example,

```
Y = (A**2)+(B**2)
  - (2.*A*B*COSF (THETA))
```

is the FORTRAN equivalent of evaluating:

$$y = a^2 + b^2 - 2ab\cos\theta$$

A single asterisk indicates multiplication, a double asterisk is exponentation, and COSF indicates that the cosine function of the argument THETA is desired.

When FORTRAN is used, the entire programme is written in FORTRAN language and fed into the machine. The computer itself then decides how to express the FORTRAN programme in its own basic code and produces a deck of cards which can be used in the future to run the problem with appropriate data.

"Interpretive systems" are also widely used. Here we substitute a simpler and more versatile language for the one built into the machine, and have the machine "translate" each instruction immediately before executing it. A typical interpretive code is the one developed at Bell Telephone Laboratories for the IBM Type 650.[1] The example, given above, now is expressed by the following series of instructions:

3	201	201	601
3	202	202	602
0	304	203	603
3	603	201	604
3	604	202	605
1	601	602	606
2	606	603	607

This appears to have little relation to the formula it is intended to represent, but we will soon see that it is entirely logical. (Although not programmed as efficiently as possible, this will serve as a suitable example.)

Each line represents one instruction and each instruction consists of 10 digits: first one digit, then three groups of three. To understand this programme it is necessary to know a little about the memory unit used by this system. It consists of one thousand locations, with "addresses" 000 to 999. In each location one 10-digit word can be stored.

Information can be sent to or from any location by specifying its address at an appropriate place in an instruction. When new information is sent to a particular location the old information is automatically erased; but information can be called for from any location without affecting the information stored there. That is, when we send information to location 601, the number previously in 601 is lost; but when we call for information from 601, the number in 601 remains there. Prior to executing the instructions given above, we will assume that a, b, and θ have been stored in locations 201, 202 and 203 respectively.

The first instruction reads

3	201	201	601

The first digit, 3, indicates that a multiplication is to be performed. The multiplier is in 201, the multiplicand is also in 201, and the product is to be stored in 601. The instruction, therefore, computes a^2 and stores the result in 601.

In a similar way, the second instruction computes b^2.

The interpretation of the third instruction is somewhat different. Since the first digit is zero, it indicates that the first group of three digits is not an address, the way it was in the previous instructions, but instead is an instruction code. In this case, 304 indicates that the cosine of the number in location 203 is to be taken and

[1] IBM Technical Newsletter No. 11.

stored in 603. This is multiplied by a in the next instruction and stored in 604; and by b in the next and stored in 605.

Finally, once we know that the code "1" is addition, and "2" is subtraction, the last instructions should be easily understood. The final result can be found in location 607.

By way of contrast, consider the programme for the same calculation using a system called SOAP on the IBM Type 650:

```
BEGIN   RAU     A
        MPY     A
        STU     ASQ
        RAU     B
        MPY     B
        STU     BSQ
        RAU     THETA
        LDD             COS
        MPY     A
        RAU     8003
        MPY     B
        STU     TEMP
        RAU     ASQ
        AUP     BSQ
        SUP     TEMP
        STU     Y       END
```

Note how much more has to be written in comparison with either FORTRAN or the Bell interpretive system. This is a very flexible programming procedure, but it is also somewhat complicated.

Without going into much detail, a few things can be pointed out. Every instruction begins with a three-letter mnemonic code. Memory locations are also given symbolic codes. The first three instructions calculate and store a^2. RAU means Reset Add Upper accumulator (part of the arithmetic unit). The effect is to clear the arithmetic unit and set up the multiplier, a. The second instruction carries out multiplication by a; the third instruction, STU, means STore Upper accummulator, and it stores the product in the symbolic location ASQ. To explain the rest of this programme is beyond the scope of this article.

Brief though our description has been, it should at least indicate that using a technique such as FORTRAN, or the Bell Laboratories interpretive code, is not beyond the capabilities of any astronomer.

3. THE ROUTINE USE OF COMPUTERS

Suppose we have become reasonably proficient programmers; what then? To what sort of problems could this technique be applied? To begin with one could automatize all the standard computations that are being done at an observatory: radial velocities, photoelectric reductions, least squares solutions, plate reductions, radio telescope reductions, statistical analyses, etc.

The process begins by making a broad outline of the computation in the form of a "flow diagram". The entire computation is divided into "boxes"; each box represents a separate phase of the problem, frequently complete within itself. The boxes

are linked by flow lines which indicate the path the computation should follow. The machine can be instructed to choose between alternate paths; for example, different procedures may be needed depending upon the sign of a result, or perhaps a particular function is best represented by one expansion for large values of the argument and another for small values. The computer can be programmed to make decisions of this kind.

The problem is usually programmed one box at a time. Each large box may be further subdivided into smaller boxes, depending upon its complexity. By using a flow diagram, the programming can be done in an orderly way and the interrelation of all the parts is clearly visible. In explaining a problem to someone else, either an astronomer or a computer expert, it is much easier to talk in terms of the flow diagram than in terms of the programme. The programme is cluttered with details of little interest to anyone but the programmer, but a properly prepared flow diagram gives the essence of the calculation.

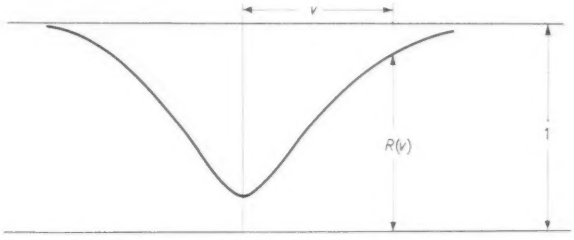


FIG. 1. Definitions for line profile computation.

Consider the following example from the theory of absorption lines. Suppose we want to find the profiles of lines formed by scattering according to the SCHUSTER-SCHWARZSCHILD model. The parameters of the problem are:

(1) the abundance factor, τ_0;
(2) the limb darkening, $I^{(0)}/I^{(1)}$; and
(3) the damping, a.

The line profile is built up point by point using the variable, v, to measure the distance from the centre of the line. The object of the calculation is to compute $R(v)$, the residual intensity at the point, v, in units of the continuum (Fig. 1).

CHANDRASEKHAR[2] has given the solution for $R(v)$, and is has been written for purposes of computation[3] as

$$R = \frac{\frac{1}{3} + Ky}{\frac{1}{3} + \tau_1 + 2y}$$

where y is tabulated as a function of $e^{-\tau_1}$, and

$$K = \frac{4}{3I^{(0)}/I^{(1)} + 2}$$

depends only on the limb darkening.

[2] CHANDRASEKHAR, S., and ELBERT, DONNA in *Ap. J.*, **115**, 269, 1952.

[3] WRUBEL, M. H. in *Ap. J.*, **119**, 51, 1954.

The quantity τ_1 depends upon a, v, and τ_0, and can be found from the theory of the absorption coefficient. It is given by the relation

$$\tau_1(a, v) = \tau_0 H(a, v).$$

HARRIS[4] has tabulated the functions $H_i(v)$, necessary to evaluate

$$H(a, v) = \sum_{i=0}^{4} a^i H_i(v).$$

We have given above everything necessary to compute profiles, but we have not put things in a logical order for computation. Obviously y must be known before we can find R, and τ_1 must be known before we can find y.

The flow diagram of Fig. 2 indicates the order in which the calculations must be carried out to compute a line profile automatically. The term "Read" means that

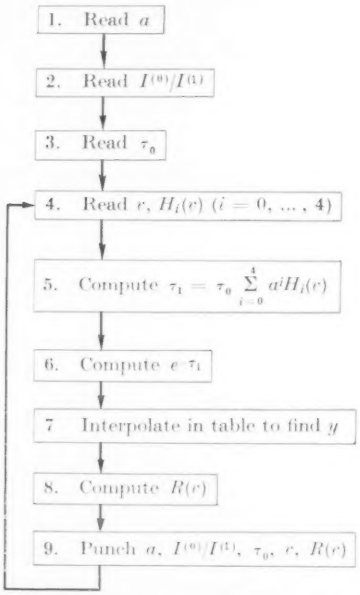

1. Read a
2. Read $I^{(0)}/I^{(1)}$
3. Read τ_0
4. Read v, $H_i(v)$ $(i = 0, \ldots, 4)$
5. Compute $\tau_1 = \tau_0 \sum_{i=0}^{4} a^i H_i(v)$
6. Compute $e^{-\tau_1}$
7 Interpolate in table to find y
8. Compute $R(v)$
9. Punch a, $I^{(0)}/I^{(1)}$, τ_0, v, $R(v)$

FIG. 2. Flow diagram for SCHUSTER-SCHWARZSCHILD model line profiles.

a card enters the input, and the information punched on the card is stored in the memory unit. (In box 4 the values of v and H_0 to H_4 are all on one card.) "Punch" means the results, and some identifying information are punched on a card by the output unit.

After the result has been calculated and punched for one value of v, the machine automatically reads the next value. This process continues until the last value of v is read. It is then possible to read new values of a, $I^{(0)}/I^{(1)}$, or τ_0, if desired.

As another example, consider the reduction of photoelectric observations. One procedure has been programmed for the IBM 650 by EDWARD C. OLSON for the National Astronomical Observatory, following a method due to HAROLD L. JOHNSON.

In order to reduce observations to the U, B, V, system, it is necessary to obtain

[4] HARRIS, D. L., III; in *Ap. J.*, **108**, 112, 1948.

the extinction of the atmosphere and the transformation of the instrumental colour system to the standard system.

For example, assume that the B-V colour is related to the instrumental yellow colour outside the atmosphere, C_y, by the linear transformation:

$$B - V = A_1 + A_2 C_y.$$

C_y is found from the raw colour, C_{y0}, by using an extinction coefficient to remove the effect of the atmosphere:

$$C_y = \frac{C_{y0} - k_1 \sec z}{1 + 0{\cdot}032 \sec z}$$

In turn, the raw colour is found from the yellow and blue deflections, Y^1 and B^1, together with the amplifier-gain step-calibrations, S_Y and S_B:

$$C_{y_0} = 2{\cdot}5 \log (Y^1/B^1) = S_B - S_Y.$$

If four or more standard stars of known B–V are observed, the values of A_1, A_2 and k_1, can be found by least squares. This will involve setting up three linear normal equations which are solved simultaneously by standard procedures. The flow diagram is given in Fig. 3.

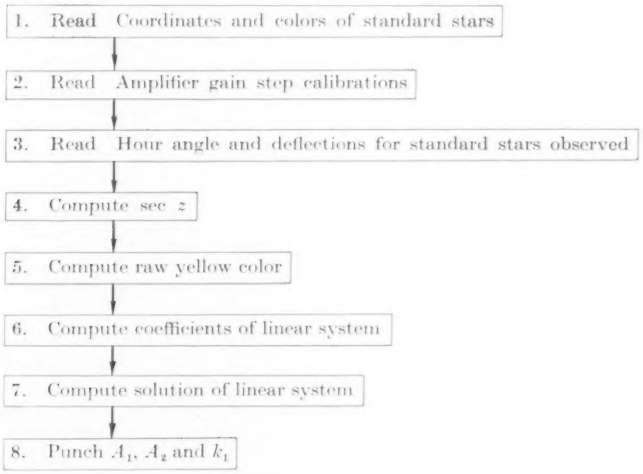

1. Read Coordinates and colors of standard stars
2. Read Amplifier gain step calibrations
3. Read Hour angle and deflections for standard stars observed
4. Compute sec z
5. Compute raw yellow color
6. Compute coefficients of linear system
7. Compute solution of linear system
8. Punch A_1, A_2 and k_1

FIG. 3. Flow diagram for obtaining B–V transformation and extinction factors.

A similar procedure is used to obtain the constants needed to go from ultra and yellow deflections to U–B and V. The individual formulae are somewhat different, but eventually the problem reduces to the solution of a system of simultaneous linear equations. This suggests that we can use the same programme for solving each linear system. Box 7 can therefore be removed from the general flow of the problem and treated separately, as a "subroutine". Whenever we need to solve a linear system, we execute the subroutine. When the solution is found, the normal flow of the computation is resumed.

A flow diagram for computing the three groups of transformation coefficients is given in Fig. 4.

The main flow is interrupted three times to execute the same subroutine—each time, of course, with different data.

I do not mean to minimize the steps that lie between the flow diagram and the final programme, but the flow diagram is half the battle. With the aid of a flow diagram, the astronomer should be able to describe his computation to the programmer. If the astronomer knows programming as well, the battle is won.

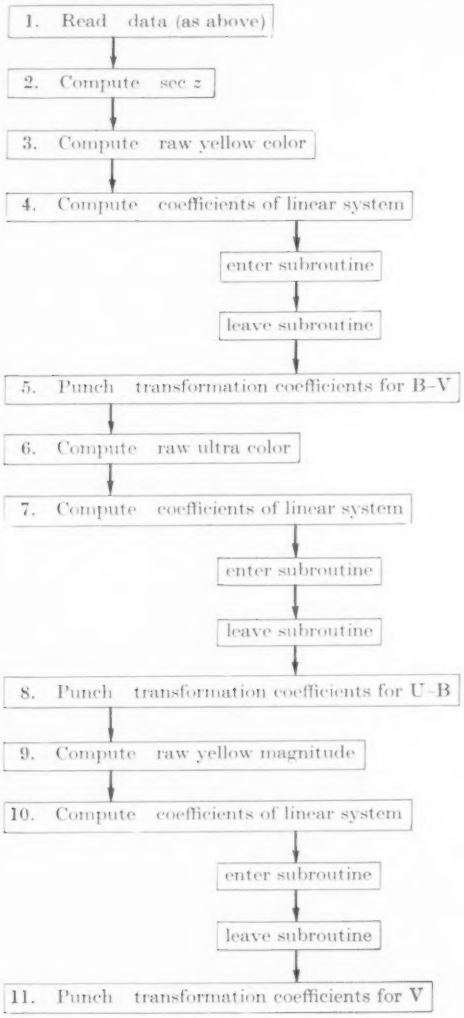

1. Read data (as above)
2. Compute sec z
3. Compute raw yellow color
4. Compute coefficients of linear system
enter subroutine
leave subroutine
5. Punch transformation coefficients for B–V
6. Compute raw ultra color
7. Compute coefficients of linear system
enter subroutine
leave subroutine
8. Punch transformation coefficients for U–B
9. Compute raw yellow magnitude
10. Compute coefficients of linear system
enter subroutine
leave subroutine
11. Punch transformation coefficients for V

FIG. 4.

4. THE IMAGINATIVE USE OF COMPUTERS

If electronic computers did nothing but the routine reductions associated with an observatory, they would be very useful. But their potential is far greater than that. As one becomes more familiar with computers, they illuminate new ways of approaching old problems; and, which is certainly more important, they bring observational

and theoretical problems that were at one time beyond our reach into the realm of possibility. For example, sequences of stellar models describing tracks of evolution in the H–R diagram require such an enormous amount of computation that high-speed machines are absolutely necessary to calculate them in a reasonable time.

To show how a simple problem can grow in generality when an electronic computer is available, consider the flow diagram of Fig. 2 again. This was a scheme for calculating line profiles according to the Schuster-Schwarzschild model. If we examine it closely, however, we see that it can be described in more general terms (Fig. 5).

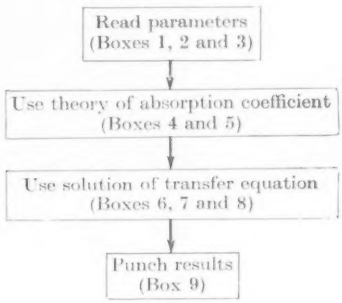


FIG. 5.

In our original application we used the results of the solution of the transfer problem for the SCHUSTER-SCHWARZSCHILD model in boxes 6, 7 and 8. We could easily substitute the solution for the MILNE-EDDINGTON model, keeping the programme for boxes 1 to 5 and box 9 intact; or we could substitute any other assumed relation between the residual intensity and the parameters.

We could further change boxes 4 and 5 if some other theory of the absorption coefficient were preferable—say, in considering the Stark broadening of hydrogen lines. In this way we may "ring the changes" on an existing programme, preserving or replacing sections as the situation demands.

Actually, electronic computers make it feasible to go beyond idealized models in which the parameters are constant with depth, to compute line profiles for model atmospheres in which ionization, damping, etc., vary with depth. Although this is possible with desk calculators as well by using weight functions or similar techniques, the ease with which these computations can be carried out on an electronic machine makes it reasonable to explore a much wider range of possibilities. For example, it would be possible to perform numerical experiments on the effect of blending or blanketing.

This flexibility of programmes is not restricted to theoretical problems. In the example we discussed concerning the reduction of observations of standard stars to obtain extinction and colour system transformations, we assumed a linear relation between the instrumental and B–V colours. This assumes the systems are not very different; but if they were, the machine could be programmed to include non-linear terms as well.

Astronomers have scarcely begun to explore the possibilities of existing computing equipment, but we might mention some of the advances that probably lie ahead.

One particularly promising application of computer techniques is in selecting, from a large memory, information which fits particular requirements. If a suitable device is perfected to transcribe a printed page directly to magnetic tape, it would become practical to record volumes of information, such as the Henry Draper Catalogue, in a form machines can easily use. This could simplify the planning of observing programmes, because it would then be possible to ask the machine to "find the co-ordinates of all A stars brighter than 10th magnitude that can be observed from McDonald Observatory in March".

Going one step further, it may one day be possible to search the literature by machine to find all references to a particular object. And, of course, there is the possibility of translation by machine, an El Dorado so many are seeking, and which will be of profit to the astronomer as well as every other scientist when found.

5. Conclusions and Recommendations

Few observatories have enough routine reductions to keep a high-speed machine busy, even for an 8-hour day. Therefore, if they do not have a theoretical department interested in machine computations they cannot justify the rental or purchase of a large calculator. Observatories connected with universities may be fortunate in having a computer facility accessible on the campus which they can share with other departments. In other cases, nearby industrial installations may be generous in offering computing time in the wee hours of the morning. (This should not disturb an astronomer.)

The success of the connection between an observatory and an electronic computer will largely depend upon the accessibility of the machine. It must be near enough so that it can be used without spending a disproportionate amount of time getting to it. It ought to be a machine for which a large programme library of standard methods is already available. (This is the case for most of the widely distributed machines and also for a few "one of a kind" machines at active research installations.) The observatory should hire a part-time programmer who would be responsible for co-ordinating the observatory's use of the machine. He would also have the responsibility of running the programmes on the machine.

Finally, at least one astronomer on the observatory staff should acquire, through experience, a knowledge of machine techniques. Otherwise, the connection between the observatory and the computer will be too remote to be effective.

The electronic computer is an important research tool. Now it is up to the astronomer to use it.

Techniques in Auroral Investigations

A. H. Jarrett

University Observatory, St. Andrews, Scotland*

Summary

The greatly increased interest shown in aurorae, chiefly as a result of the International Geophysical Year, and the techniques associated with auroral research are discussed.

It is perhaps appropriate to commence with a quotation from a recent article by Professor S. Chapman (1956), stating that "the aurora is the only visible manifestation on the Earth of a complicated series of phenomena that originate from disturbances in the internal and upper layers of the Sun".

1. Visual and Photographic Auroral Observations

Thousands of photographs of this beautiful phenomenon have been amassed over the past fifty years; from their analysis several important features have emerged. Perhaps the most interesting one is that the lower regions of almost all auroral arcs occur at an altitude of nearly 100 km. This type of measurement is only possible by parallactic photography, i.e. by photographs taken simultaneously from two or more observing stations.

The Norwegians, under the leadership of the late Professor C. Störmer have been prominent in this field, and in the British Isles a research group has been working for many years under the direction of J. Paton.

Recent publicity due to the International Geophysical Year has drawn much attention to the scheme for visual observations of aurorae, organized at Edinburgh by Paton. Briefly, the collected results of numerous enthusiastic, voluntary observers are examined and plotted on specially prepared maps, so as to indicate the position and extent of the various auroral forms over the Earth's surface. Paton has laid particular stress on obtaining observations from ships and aircraft on the North Atlantic routes, in order to ensure as complete a coverage as possible of the northern regions. It will be of great interest to examine these records over a complete solar cycle.

An extension of Paton's scheme has been the installing of several All-Sky Cameras at selected sites throughout Scandinavia, one in Iceland and one at St. Andrews. These cameras were constructed at Uppsala; they photograph the entire sky every minute (with an exposure of 25 seconds on 16 mm cine-film) throughout the hours of darkness. Time-markings are printed on the film at appropriate intervals. A grid consisting of a set of small lamps aligned in a N–S and E–W direction above the camera, and ranging in elevation from the horizon to the zenith, is recorded on each frame. Consequently the bearing and elevation of any auroral feature can be determined directly. Cameras of a similar type are operating in North America and also in the Antarctic.

* Now at the Queen's University of Belfast, N. Ireland.

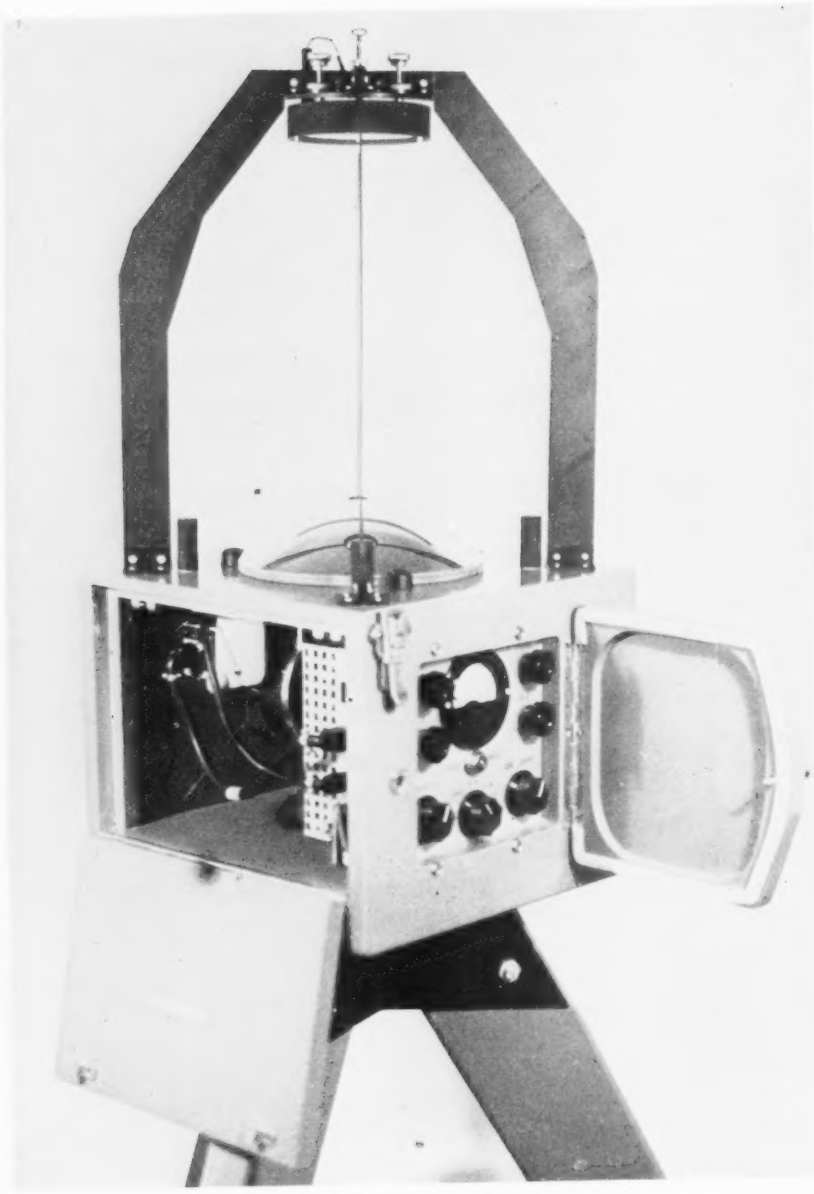


FIG. 1. All-Sky Camera, St. Andrews.

The cine-camera is in the box underneath the convex mirror. The image of the sky is photographed in the plane mirror held by the metal supports.

Fig. 2. All-Sky Camera, St. Andrews.

The camera can be seen mounted underneath the perspex dome at the top of the tower. The dome is continuously fed with hot air to prevent condensation during the night.

It does not seem to be generally realized that an ideal site for general auroral observations should be several degrees of geomagnetic latitude to the south of the zero of maximum occurrence, and *vice-versa* in the southern hemisphere. VESTINE (1944) gives 65°N geomagnetic latitude as the zone in the northern hemisphere. St. Andrews, for example, is favourably placed at 59°N geomagnetic latitude.

Colour photography of the aurora has quite interesting possibilities, particularly for determining the relative intensities of the auroral radiations. HARANG (1933) took

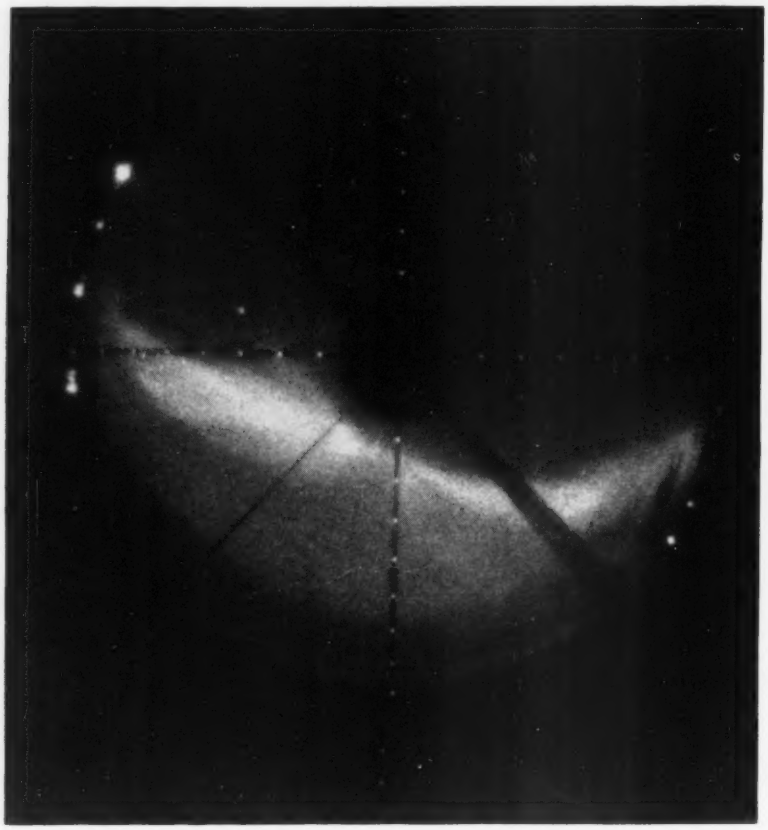


Fig. 3.

All-Sky Camera photograph, taken in the night of 1957 September 29–30, showing auroral activity reaching to 70° elevation above the southern horizon. During this intense display the activity as seen from St. Andrews moved from the North through the zenith into the southern part of the sky. (North is on top, East to the left.)

filter photographs, simultaneously in the green and violet, of various auroral forms and obtained isophotes for some bands and draperies. Recent developments in high-speed colour-film make possible a considerable extension of this work. At the time of writing the only visual photometer for auroral colour measurements known to the author is the one used by GADSDEN (1957) in his work on the colour of the zenith twilight-sky.

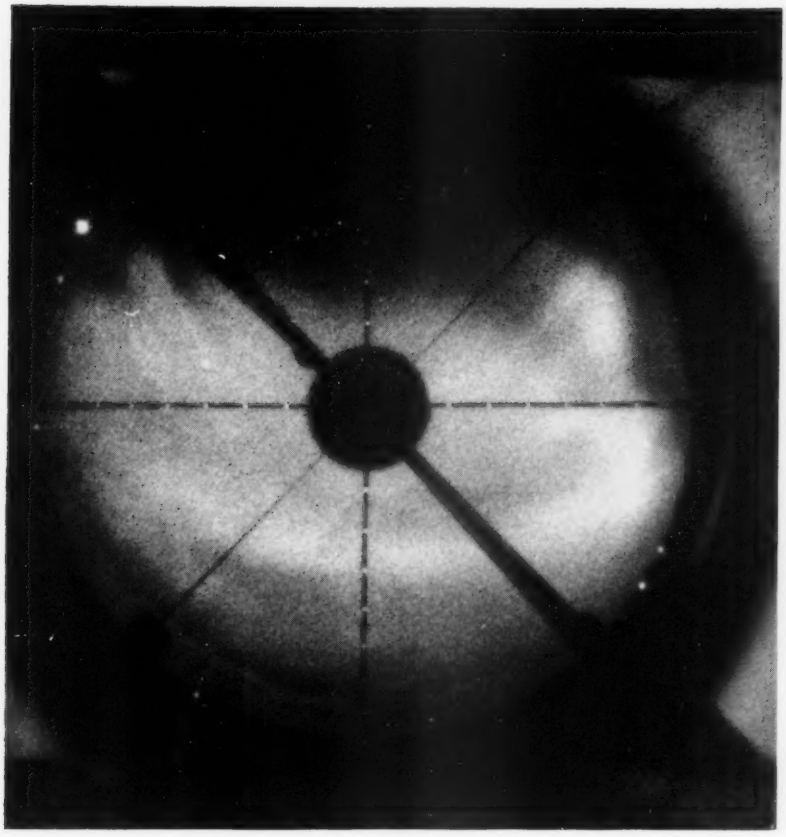


Fig. 4.
This is a later stage of the great display of 1957 September 29–30. The activity was particularly intense in the southern part of the sky. (North is on top, East to the left.)

2. The Fabry-Perot Interferometer in Auroral Research

The earlier experiments by BABCOCK (1923) on the 5577 Å night-sky line, and by VEGARD and HARANG (1934) on the same green auroral line, using a Fabry-Perot interferometer and ordinary dye-type filters have been reviewed with renewed interest in the light of the rapid progress in multilayer coatings since 1949. Several trials by the writer JARRETT (1955) on auroral glows have recorded Fabry-Perot fringes from the 5577 radiation with exposures of as little as 20 minutes, using a multilayer interferometer placed in front of an f/2 camera. The objectives of these experiments are the temperature distribution in the aurora from measurements of the fringe halfwidths, and line-of-sight motions of the aurora from displacements of the fringes. Further possibilities arise from the narrow bandwidth ($\simeq 40$ Å at 5200 Å and 75 per cent transmission) all-dielectric interference filters, as reported by RING (1956); of particular interest would be to know the extent of the hydrogen Balmer radiation in a given display. Moreover, by the relative enhancement of the auroral radiation, these filters should greatly increase the possibility of observations during moonlight.

The importance of the presence of Hα and Hβ in some of the spectra, taken by VEGARD (1940, 1950) and later by MEINEL (1951) is that this is an indication of showers of hydrogen occasionally entering the Earth's atmosphere.

The wider-bandwidth ($\simeq 80$ Å at 5500 Å) interference filters are a valuable aid for visual observers of aurorae, especially when cloud cover makes for difficulty in deciding whether activity is really present, or whether it is just scattered light from street lamps. A very convenient arrangement is to have them made into a pair of goggles.

3. AURORAL SPECTROSCOPY

Turning now to another topic of auroral research, namely spectroscopy, we are at once up against the difficulty arising from the general low intensity of the activity. At its best it is seldom brighter than the light from the full moon. Clearly, a spectrograph of high light-gathering power is essential. Since the aurora is effectively a luminous surface of appreciable angular dimensions, the light-gathering power of the spectrograph is fixed by the speed of the camera. The tendency is to use a high-speed camera together with a strongly dispersive element.

The slit-type spectrograph used by VEGARD (1933) in his earlier work had an f/2·5 camera and a two-prism dispersing medium, giving a dispersion of about 280–270 Å/mm in the 4000–6000 Å region. To record the fainter auroral lines with such an instrument requires an exposure of many hours. In fact, the procedure has been to expose on several successive aurorae occurring over a whole winter season. This spectrum consequently represents an average intensity distribution for the whole auroral activity during that period.

Since 1949 several faster auroral spectrographs have been devised. Noteworthy is, e.g. that of f/1·2 installed at Tromsö by VEGARD and TONSBERG (1952). The dispersion of this prism-instrument is 41 Å/mm at 4000 Å, 100 Å/mm at 5000 Å, and 240 Å/mm at 6500 Å. The collimator- and camera-lens have an aperture of 178 mm, and the two prisms are of corresponding dimensions. In 1951 a faster spectrograph (f/0·65) was also mounted at Tromsö. Details are given in a paper by COJAN (1947). Unfortunately, the dispersion is very inferior to the earlier equipment, being only 224 Å/mm at 5000 Å, and 515 Å/mm at 6500 Å.

The most powerful slit-type spectrographs in use for auroral research at the present time are those constructed for the Cambridge Research Centre of the United States Air Force. They are grating-type instruments with a mirror collimator and an f/0·8 Schmidt camera, and are developments of a spectrographic camera designed and constructed by MEINEL (1954) at the Lick Observatory in 1948. The effective focal ratio is 1 : 1·0, due to masking by the grating and plate-holder. The green auroral line is often recorded in 10 minutes or less even from faint glows. The dispersion in the third order is 40 Å/mm. These spectrographs have been set up at selected sites (in North America, one at Invercargill in New Zealand, and another at St. Andrews) for participation in the current International Geophysical Year programme. The chief aim of these I.G.Y. spectrographic proposals is to extend the study of the variations of the wavelengths and intensities of the atomic lines and molecular bands with altitude and type of aurora. Especially emphasized is the need for further observations of the hydrogen lines. In addition, more temperature determinations of the auroral region should be made following the method of VEGARD and TONSBERG (1952), using the intensity distribution of the rotational

components of the R-branch of the negative nitrogen bands. It is interesting to note that VEGARD concludes from his work that there is no indication of an increase in temperature with increasing altitude. This conclusion is at variance with other researches, for example, with the temperatures measured during the recent experiments with rockets.

With auroral features of short duration, such as a rayed structure which rarely remains in one position for more than a few minutes, there is little point in imaging a selected portion onto the spectrograph slit; for, unless the activity is exceptionally bright, there is no likelihood of recording it in the time available. It is better to point the instrument directly at the display and to select via the collimator slit a particular region. Used in this manner, the St. Andrews spectrograph integrates the light from a region nearly 7 degrees in extent.

Every effort should be made to cut down light-losses in the spectrograph. For instance, it is unwise to introduce the light via a slit and to reflect it then through a right angle onto the main axis of the collimator mirror; (a mirror is used in the collimator, rather than a lens, to minimize absorption losses). It is more efficient to mount the slit off-axis, using a long focal-length collimating mirror. One should assume that the corrector plate and the field-flattener surfaces and dispersing prism-faces are anti-reflection coated wherever appropriate. In the case of a grating it will of course be blazed for a particular order.

The brighter auroral lines and bands (e.g., 3914, 4278, 5577 and 6300 Å) are suitable for recording with a slitless spectrograph. LEBEDINSKY (1956) has found that sufficiently dense images of these lines are given, in the cases of bright aurorae, with exposures of about one minute. These exposures are about ten times shorter than those necessary for the same lines using slit-spectrographs of identical focal-ratio and photographic emulsion.

4. PHOTOMULTIPLIER APPLICATIONS

Over the post-war years there has been considerable improvement in photomultiplier tubes with the result that it is now possible to scan the spectra photoelectrically. Such a photoelectric scanning spectrograph has been constructed by HUNTEN (1953); this type of instrument has the great advantage of giving a direct trace of the spectrum (if the output is fed into a pen recorder) within a few seconds, depending naturally on the scanning rate.

ARMSTRONG (1956) has arranged a photomultiplier to scan Fabry-Perot interference fringes from airglow and aurorae. In his equipment the air-pressure in a cylinder containing the interferometer was raised, and the fringe profile obtained by plotting the photomultiplier current against pressure as the air was allowed to leak away. A later version used a photomultiplier with a pulse-counting circuit. The conclusion is, that for quiet aurorae the photoelectric interferometer is quite promising. KARANDIKAR (1956) arrived at a similar conclusion as to the success of this method for temperature measurements over stationary regions of an auroral display, or for rapid surveys across a homogeneous quiet arc and other auroral forms.

5. OBSERVATIONS OF DAYTIME AURORAE

There are no records of aurorae being seen during the daytime, other than during the evening- and dawn-twilight periods. CHAPMAN and STIBBS (1953) discussed the possibility of observing daytime aurorae during a total solar eclipse, the

path of totality crossing the auroral zone. An aircraft flight (PATON, 1954) made on June 30, 1954 to a position within the total eclipse belt south-west of Iceland, at an altitude of 8500 feet, failed in the sighting of any auroral activity by the observers aboard. If an aurora was actually present at the time, it probably faded into the sky-background which was estimated to have a brightness certainly greater than that of the full moon.

The possibility of observing daytime aurorae from high-flying aircraft came recently under consideration at St. Andrews (BYARD and JARRETT, 1959) but there seems little chance of success unless a height of over 30 kilometres can be reached.

However, daytime auroral activity does not escape detection completely, for continuously operating radio-echo equipment shows that it certainly does occur but that it is usually weak. In this connection it is reasonable to suppose that the emissions from daytime aurorae may differ from those at night, due to different states of the Earth's atmosphere when subject to solar radiation and when the Sun is absent.

6. RADIO-ECHO INVESTIGATIONS

Most radio-echo information about the auroral regions has been achieved on metre-wavelengths. The earliest experiments were made by HARANG and STOFFREGEN (1940) on a frequency of 40 Mc/s, using a pulse transmitter of peak power 4 kilowatts. LOVELL, CLEGG and ELLYETT (1947) reported echoes on 46 Mc/s, using equipment designed for the detection of meteors. Various other workers in Scandinavia and North America have since found echoes on frequencies within the range 30 to 150 Mc/s. A concise account of work already carried out in this field is given in STÖRMER's (1955) book *The Polar Aurora*. Attention is drawn to the proposal by HARANG and LANDMARK (1954) that the echoes cannot be due to direct backward scattering, but may be explained by assuming scatter from land or sea via the E_s-layer. This idea is based on the observation that echoes appear at smaller ranges on 35 Mc/s than on 74 Mc/s. However, BULLOUGH and KAISER (1954) at Jodrell Bank, as well as MCNAMARA and CURRIE (1954) at Saskatoon, find that this proposal does not fit in with the observed results. The radio-echo techniques, besides giving a value for the range of the activity, also give an indication of the electron density at auroral heights. For detailed accounts of radio-wave reflections from aurorae reference should be made to the papers by KAISER (1956) and FORSYTH (1954).

FORSYTH, PETRIE and CURRIE (1949) have reported the reception of radio noise at a frequency of 3000 Mc/s. So far this noise has not been verified by other workers. Presumably, its strength is a function of the intensity and the type of display.

In conclusion, we may emphasize that a vast amount of experimental evidence is still needed, both on the optical- and the radio-side of the problems, in order to arrive at a satisfactory explanation of many of the varied auroral features.

REFERENCES

ARMSTRONG, E. B.	1956	*The Airglow and Aurorae*, edited by E. B. Armstrong and A. Dalgarno p. 366, Pergamon Press, London.
BABCOCK, H. D.	1923	*Ap. J.*, **57**, 209.
BULLOUGH, K. and KAISER, T. R. .	1954	*J. atmos. terrest. Phys.*, **5**, 189.
BYARD, P. L. and JARRETT, A. H. .	1959	*J. atmos. terrest. Phys.* (in preparation).
CHAPMAN, S. and STIBBS, D. W. N. .	1953	*Sky and Telescope*, **12**, 35.

CHAPMAN, S.	1956	*International Geophysical Year Instruction Manual*, Part II, p. 89, *Aurora and Airglow*, Pergamon Press, London.
COJAN, J.	1947	*Ann. Astrophys.*, **10**, 33.
FORSYTH, P. A.	1954	*Proceedings of Conference on Auroral Physics* in July 1954; Geophysical Research Paper No. 30, p. 117; Air Force Cambridge Research Centre, Cambridge, Mass.
FORSYTH, P. A., PETRIE, W. and CURRIE, B. W.	1949	*Nature* (London), **164**, 453.
GADSDEN, M.	1957	*J. atmos. terrest. Phys.*, **10**, 176.
HARANG, L.	1933	*Geofys. Publ. Oslo*, **10**, No. 8.
HARANG, L. and STOFFREGEN, W.	1940	*Hochfrequenztechnik und Elektroakustik*, **55**, 105.
HARANG, L. and LANDMARK, B.	1954	*J. atmos. terrest. Phys.*, **4**, 322.
HUNTEN, D. M.	1953	*Canadian J. Phys.*, **31**, 681.
JARRETT, A. H.	1955	*Quart. J. Roy. Meteorol. Soc.*, **81**, No. 347, 106.
KAISER, T. R.	1956	*The Airglow and Aurorae*, p. 156; Pergamon Press, London.
KARANDIKAR, R. V.	1956	*The Airglow and Aurorae*, p. 374; Pergamon Press, London.
LOVELL, A. C. B., CLEGG, J. A. and ELLYETT, C. D.	1947	*Nature* (London), **160**, 372.
LEBEDINSKY, A. I.	1956	*The Airglow and Aurorae*, edited by E. B. Armstrong and A. Dalgarno, p. 388; Pergamon Press, London.
MCNAMARA, A. G. and CURRIE, B. W.	1954	*J. Geophys. Res.*, **59**, 279.
MEINEL, A. B.	1951	*Ap. J.*, **113**, 50.
	1954	*Mitt. Hamburger Sternwarte in Bergedorf*, Band 23, No. 255, pp. 91–99.
PATON, J.	1954	*J. Brit. Astron. Assoc.*, **64**, 376.
RING, J.	1956	*Astron. Contrib.*, University Manchester, Series III, No. 39.
STÖRMER, C.	1955	*The Polar Aurora*, Clarendon Press, Oxford; p. 183.
VEGARD, L.	1933	*Geofys. Publ. Oslo*, **10**, No. 4.
	1940	*Terr. Magn. Atm. Elect.* **45**, 5.
	1950	*Nature* (London), **165**, 1012.
VEGARD, L. and HARANG, L.	1934	*Geofys. Publ. Oslo*, **11**, No. 1.
VEGARD, L. and TONSBERG, E.	1952	*Geofys. Publ. Oslo*, **18**, No. 8.
VESTINE, E. H.	1944	*Terr. Magn. Atm. Elect.* **49**, 77.

Note on Optical Observations of the Russian Satellites

D. W. Beggs

The Observatories, Cambridge, England.

The appearance of the first two Russian satellites in orbits inclined at 65° to the equator resulted in the need for optical observations at higher latitudes than had been expected. No teams similar to those of the American "Moonwatch" had been organized in the United Kingdom, and it is not surprising in view of the unfavourable weather conditions at the time that theere were no confirmed observations of the 23-inch diameter sphere of Sputuik I in this country, although a few sightings were obtained elsewhere. A single observer, if not aided by accurate predictions, has little chance of detecting satellites below naked-eye visibility. However, Sputnik II and the carrier-rocket of Sputnik I, both being naked-eye objects, greatly reduced the need for specialised and expensive equipment for accurate positional observations.

Positions to a few seconds of arc are fairly easy to obtain for objects as bright as these. Times to 0·01 seconds likewise present no particular difficulties. It is doubtful however, whether it is necessary to aim at such an accuracy in the case of non-spherical satellites which encounter appreciable air-resistance near perigee. The erratic rate of decrease in the period of Sputnik II was probably due in the main to the varying cross-sectional area presented by the satellite as it tumbled in its orbit; (see, e.g., D. G. King-Hele in *Nature*, **181**, 738, 1958, and L. G. Jacchia in *Sky and Telescope*, **17**, 278, 1958). In particular during the final days of a satellite's life, before it enters the dense layers of the Earth's atmosphere, we require observations which can be reduced rapidly, rather than data of the highest accuracy.

Worthwhile observations can be made with the aid of simple instruments such as normal hand-cameras, provided these have f-ratios of at least 4·5. The attached photograph of Sputnik II (Fig. 1), transiting beneath the North Pole, shows the type of observations obtained by this means. The interruptions were made to provide timing marks, the segments being of 4·1, 8·2 and 9·3 seconds duration, respectively. The variation in the intensity of the trail was due to the satellite's rotation.

Timing can be the major difficulty for amateurs, for whom this photographic method is well suited, but a carefully rated stop-watch compared with radio time-signals enables accuracies between a few tenths of a second and one second to be achieved. Positional accuracy is of the order of a few minutes of arc. Photographic methods, however, are limited to times when the sky-background is dark enough, or free from cloud, to provide a sufficient number of stars for reference purposes. Theodolite-type instruments, provided they have adequate fields of the order of a few degrees, are more versatile; observations of large satellites can then easily be obtained in bright twilight-conditions and even when there are fairly extensive clouds or thick haze as so often occurs in the British Isles. This type of instrument, in contrast e.g. to "Moonwatch"-type telescopes, enables a single observer to choose his point of observation according to conditions; he is not restricted to a particular

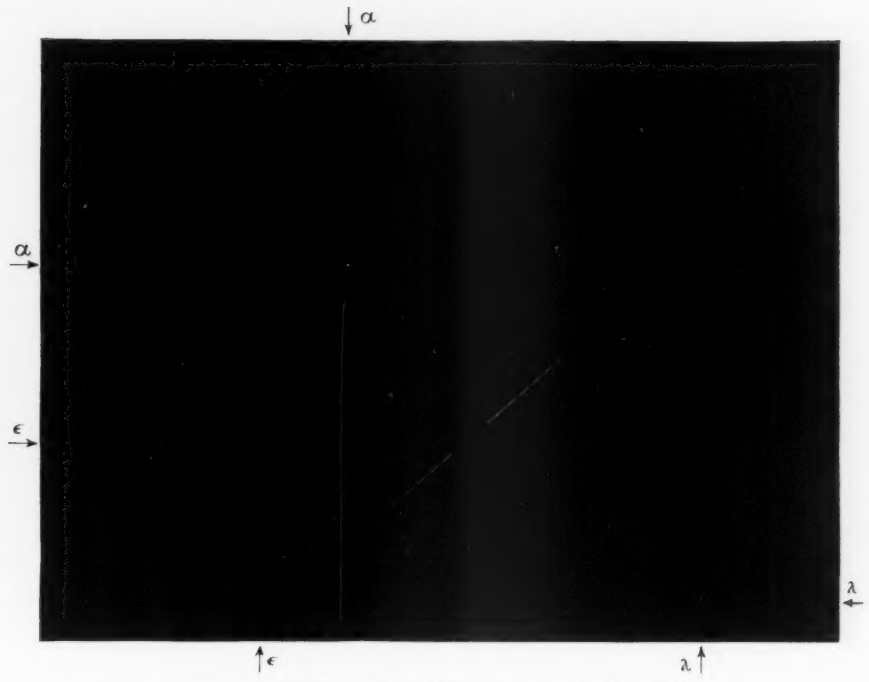


Fig. 1. SPUTNIK II: 1958 January 19, 18^h32^m U.T.
Three reference stars are marked : $\alpha = \alpha$ UMi, $\epsilon = \epsilon$ UMi, $\lambda = \lambda$ Dra. (Vito-B camera, f/3·5, on Ilford HPS-film; D. W. Beggs, Observatories, Cambridge.)

azimuth. With the aid of a chronograph it is possible to obtain two or more observations of a particular passage of a satellite. The accuracy achieved with simple equipment of this kind is again of the order of a few minutes of arc in position and a few tenths of a second in time. A series of observations by this method were obtained at Cambridge, and the results sent to the various centres collecting artificial-satellite data.

The Two Bright Comets of 1957

I

Arend-Roland (1956h) and Mrkos (1957d)

J. G. Porter

Royal Greenwich Observatory, Herstmonceux

Not since 1910 have observers in northern latitudes had an opportunity of studying a really bright comet, and the appearance of two such bodies in 1957 aroused considerable interest.

The first of these comets was discovered by S. Arend and G. Roland at Uccle on the night of 1956 November 8 in the course of routine work on asteroids. The image of the comet, then of the tenth magnitude, was found on two Kodak 103aO plates exposed in the twin astrograph. Preliminary orbits showed that Comet Arend–Roland (1956h) was approaching perihelion and would grow brighter; it moved south until the beginning of April, but its appearance in the southern hemisphere was something of a disappointment. However, as it turned north and became visible in northern latitudes towards the end of April, it came fully up to expectations as soon as it was possible to see it against a really dark sky and well clear of the horizon. By the night of April 23–24 the main tail was 30° long, and the comet was visible to the naked eye as a circumpolar object for about a week after this. The comet was followed telescopically for almost a year, the last recorded observation being made at Flagstaff on 1958 April 11, when the image was of magnitude 21 (Roemer, 1958). The motion of the comet was undoubtedly hyperbolic; the best orbit available is that of Haségawa, (1957), which is based on 116 observations in 7 normals, covering the period 18 November 1956 to 28 June 1957:

T 1957 April 8·03117 U.T.

ω	308°·78080		q	0·316062
Ω	215°·15909	1950·0	e	1·0002297
i	119°·94946			

The most remarkable feature of this comet was the brilliant "spike" which projected from the head in the sunward direction. It was seen as early as 22 April, and was plainly visible to the naked eye on April 25, estimates of its length at that time varying from 7° to 15°, according to observing conditions; on April 25·8 it was directed almost exactly in the prolonged direction of the main tail and was sharply defined on both sides. Before this date, the spike was inclined to the right, while

on later dates it was inclined to the left, so that in all it appeared to have swung through nearly 20° relative to the radius vector. In all cases one side remained sharply limited, but on the other side a diffuse glow filled the obtuse angle between the main and secondary tails. (See Fig. 1 and Fig. 2).

Fig. 1. Comet Arend-Roland: 1957 April 24·9. Exposure time 36 minutes: 6 in. Cooke Triplet, 26 in. focal length, Observatory Ascot; Kodak Oa–0 plate; Aviol 2 Ilford filter. (Photo: Dr. R. L. WATERFIELD.)

Explanations of this anomalous tail were given in various quarters (WHIPPLE 1957, LARSSON-LEANDER 1957, ARMSTRONG 1957); it will suffice to quote here that of WHIPPLE: "The sunward tail must almost certainly have resulted from the concentration of cometary debris over an area in the orbital plane. Seen at moderate angles to the plane, the material possessed too low a surface brightness to be easily observed, but seen edge-on it presented a concentrated line of considerable intensity."

There is certainly nothing new in the appearance of such a sunward tail, for "beards" as they are more commonly called were observed in the comets of 1823, 1851 IV, 1862 II, 1877 II, 1882 II among others; some of these comets were the subject of detailed investigations by BREDICHIN (1875) who gave an explanation somewhat similar to that of WHIPPLE. Thus, in speaking of the anomalous tail of the comet of 1823 he says: "Les particules de cette queue décrivaient donc la même (presque) orbite que le noyau, et cet appendice peut être regardé comme une traineé des météores, dont le plus grand était le noyau."

Fig. 2. Comet Arend-Roland: 1957 April 25. Exposure time 1 hr; Bruce Telescope, Sternwarte Heidelberg; Observer: NECKEL. (Photo: Prof. H. KIENLE.)

These conditions are illustrated in the following table, which gives the geocentric distance Δ of the comet, the distance b of the Earth above the plane of the comet's orbit (in astronomical units), the angle F which the line of sight made with that plane, and the angle Earth–Comet–Sun, represented by E.

	Δ	b	F	E
April 23	0·582	+0·04084	+4°·0	129°
24	0·594	+0·02604	+2°·5	122°
25	0·608	+0·01123	+1°·1	116°
26	0·625	−0·00358	−0°·3	111°
27	0·644	−0·01839	−1°·6	106°

The date of passage of the Earth through the plane of the orbit was April 25^d $18^h{\cdot}2$ and it is clear that the changes in the degree of foreshortening, as shown by the angle F, support the explanation given by WHIPPLE; but any attempt to calculate the angle which the spike formed with the radius vector must lead to doubtful results. No single observation of the tail could determine its orientation, even under ideal conditions, without making certain assumptions. Thus if the formulae adopted by BESSEL (1836) in his investigation of HALLEY's comet are used, the assumption is implicitly made that the tail lies in the plane of the orbit. With such an amount of foreshortening, however, the slightest deviation from that plane, or the smallest errors in determining position angles would entirely vitiate the results of such calculations. The most that we can conclude from the observations is, as stated by WHIPPLE, that the anomalous tail was due to a thin sheet of dust spread out in the plane of the orbit, many millions of miles in breadth but only about 10,000 miles or less in thickness, and distributed behind the comet, on the outer side of the orbit.

The geometry of a comet's tail is by no means the simple matter described in elementary text-books. It is clear that forces other than radiation pressure are at work, and although WHIPPLE's icy-conglomerate theory for comets is capable of providing an explanation, it would seem from a study of other recent comets (OSTERBROCK, 1958) that the intersection with interplanetary matter of material expelled from the head must also be taken into account.

The second bright comet, 1957d, was announced by Antonin MRKOS on 2 August, and was first seen by him with the naked eye while measuring the night-sky glow at the mountain observatory Lomnicky Stit, in Czechoslovakia. A number of independent discoveries were announced; these included those of KURAGANO at Yokohama (July 29), Peter CHERBAK, an American airlines pilot (July 31) and Clive HARE, a young amateur of Tamworth, England (August 3), but for practical reasons, the name Mrkos is alone retained for this comet. At the time of discovery the comet was near perihelion, and was of the second magnitude. It soon moved out from the Sun and for several weeks became a conspicuous object in the north-west after sunset, with a tail more than 5° in length. The comet had faded from naked-eye visibility by the end of September, and by the end of October was too near the sun for further observation. It was recovered at the end of January 1958, and continued under observation until 1958 July 9, when it was photographed at Flagstaff as a diffuse object of magnitude 19·0.

In its early stages the comet had two distinct tails, of which the brighter (visually) was decidedly curved. By August 21 this tail had faded considerably, and the straight tail had become brighter and displayed numerous knots of material in rapid motion. The extraordinary changes that took place in this tail are particularly well shown in the photographs taken with the Schmidt telescope at the Observatory of Haute-Provence (DUFAY 1958).

Like Comet Arend-Roland this comet had retrograde motion, in a highly inclined orbit with small perihelion distance, but the planes of the two orbits were inclined to each other at about 134° and there is no real similarity between the two comets.

The best orbit of this comet available is that of CANDY (1957), based on 19 obser- grouped in three normals 1957 August 4·0, 8 0 and 12·0.

T 1957 August 1·4320 E.T.

ω	40°·274		q	0·35511
Ω	67°·651	1957·0	e	1·0
i	93°·816			

Although these parabolic elements sufficed to provide an ephemeris, there is a residual of 11″ in the middle declination, so that the true motion of the comet may well depart from parabolic. A definitive orbit is to be computed by SCHRUTKA.

Although the unexpected appearance of these two comets provided a wealth of new observational material, it is perhaps of even greater importance that they drew attention to the fact that the solar system still has many unsolved problems and that far too little attention is paid to these. Only a mere handful of observers are interested in cometary research, and these are mainly in the northern hemisphere; if Comets Arend-Roland and Mrkos have done something to revive an interest in comets, they will achieve an added importance in the history of astronomy.

REFERENCES

ARMSTRONG, E. B.	1957	*Observatory*, **77**, 135.
BESSEL, W.	1836	*Astron. Nachrichten*, **13**, 300–302.
BREDICHIN, T.	1875	*Ann. Obs. Moscow*, **5**, 60.
CANDY, M. P.	1957	*U.A.I. Circ.*, No. 1615.
DUFAY, M.	1958	*L'Astronomie*, p. 258. (June).
HASÉGAWA, I.	1957	*U.A.I. Circ.*, No. 1619.
LARSSON-LEANDER, G.	1957	*Observatory*, **77**, 132.
OSTERBROCK, D. E.	1958	*Ap. J.*, **128**, 95
ROEMER, E.	1958	*P.A.S.P.*, **70**, 415.
WHIPPLE, F. L.	1957	*Nature (London)*, **179**, 1240.
	1957	*Sky and Telescope*, **16**, 426.

II

Cometary Physics and the Comets Arend-Roland and Mrkos

N. RICHTER

Sternwarte Sonneberg, Germany

DURING recent years many astronomers have expressed the fear that the "great comets" have died out. Since the return of HALLEY's comet in 1910 astronomers have been waiting in vain for an object of similar importance in the knowledge that the previous century had brought us no less than five of these. Comets Arend-Roland (1956h) and Mrkos (1957d) appeared just at the right moment to prove that great comets still exist and the statisticians can state with satisfaction that their frequency corresponds exactly to that during the last century.

For the modern astrophysicist these comets were a most welcome and long-awaited event. True, the physics of the comets has made appreciable progress during the past twenty years. New thoughts and theories concerning the structure of cometary nuclei have been worked out, for example WHIPPLE's model[1]. The rôle played by the selective radiation pressure and the solar corpuscular radiation in the formation of cometary tails (BIERMANN's theory[2]) provided new points of view for the understanding of the previously known relationships between comets and the Sun. Considering a comet as a plasma, that is to say, as a mixture of ions and free electrons, as well as solid particles, gave quite new aspects and it has been predicted that large comets may produce radio radiations. The identification of cometary emissions and the explanation of their spectra has been documented in the atlas published by SWINGS and HASER[3]: nevertheless it appeared that further progress, in spite of the outstanding accomplishments of laboratory spectroscopy, were not to be expected until cometary spectra were obtained with appreciably higher resolution than had been possible hitherto. The proportion of dust in cometary heads and tails could only be determined by extensive polarization measurements. For these and many other new problems it was necessary to have a large comet as a test-object. The improvement of astrophysical instruments has made possible many types of observations that could not be made on the great comets during the last century. Modern spectrographs on powerful telescopes are now available at many observatories. Photoelectric observations have been developed to high precision. Instruments for the observation of polarization can be combined with photometers, monochromators, and interference filters in order to obtain photometric data about the different components of the light of comets. Radio telescopes give additional data. Thus astrophysicists were well equipped to test new methods on the two large comets of 1957 and to compare the results with the properties of modern cometary models. It was a great advantage that they could prepare for the observation of Comet Arend-Roland with particular care. It was discovered on 8 November 1956, and they had thus nearly six months before its perihelion passage in April 1957. Such long notice has not been available since the return of HALLEY's comet in 1910. A period of fine weather in April and May 1957 enabled the collection of an enormous amount of observational material throughout the world. Although it will take a considerable time until all these data have been analysed it is nevertheless already possible to say something about the results achieved.

Comet Arend-Roland brought some surprises. The first unusual feature was the appearance of a spike-like auxiliary tail, directed towards the Sun, such as had never before been observed so conspicuously. The illustration of the orbital position, Fig. 3, shows that the whole phenomenon is a perspective effect caused by a fan-like auxiliary tail which forms an angle of some 100–120° with the main tail. When the Earth passed through the plane of the orbit of the comet on the evening of 25 April 1957 the angle had increased to 180°, and the wide fan of the secondary tail, situated in the orbit of the comet appeared, seen edge-on, as a luminous spike. It is interesting that at the time of its greatest extent it was about $5{\cdot}6 \cdot 10^7$ km long, and decidedly longer than the main tail which at the time was only $4{\cdot}5 \cdot 10^7$ km.

[1] WHIPPLE, F. L., *Ap. J.*, **111**, 375, 1950; *Ap. J.*, **113**, 464, 1951; *Ap. J.*, **121**, 750, 1955.

[2] BIERMAN, L., *Z. Astrophys.*, **29**, 274, 1951. *Z. Naturforschung*, **7a**, 127, 1952.

[3] SWINGS P. and HASER, L., *Atlas of representative cometary spectra*, Institut d'Astrophysique, Liège, 1956.

A further surprise was provided by the spectrum of the comet. It showed an extremely strong continuous spectrum, the total brightness of which exceeded more than three times the integrated intensity of the cometary emissions. This strong continuous spectrum was characteristic not only of the head of the comet, but also occurred in the main and the auxiliary tail together with the typical emissions of CO^+ and N_2^+. The sodium line appeared much stronger than had previously been observed at equal distance ($r = 0{\cdot}60$) from the Sun, and showed marked fluctuations. This line was responsible for the yellowish colour of the comet during the last days of April. This strong continuum made it appear that the cometary head must contain a high proportion of solid, probably dust-like particles. This view was strengthened

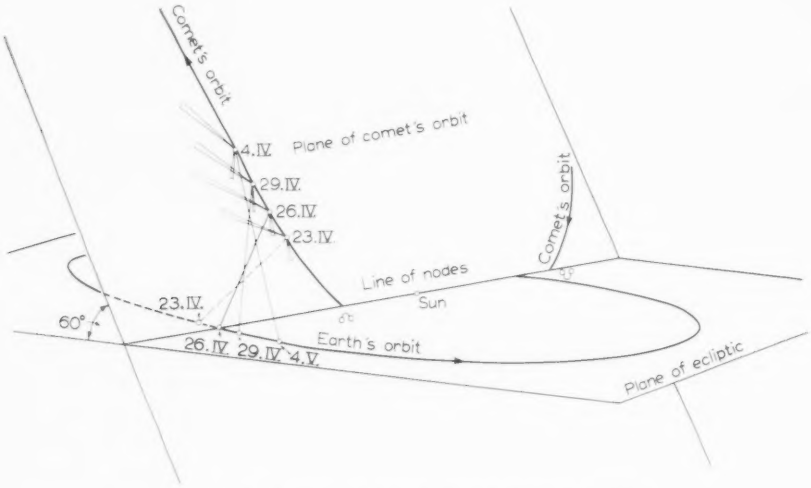


Fig. 3. The orbit of Comet Arend-Roland (1956h)

by the high values, up to 30 per cent, found by several observers for the polarization. The relationship between the degree of observed polarization and the phase angle agreed with that found for meteorites.

It was with greatest satisfaction that observers were able to establish that there were radio emissions from the comet—as had been predicted. They were found at a wavelength of 11 m (KRAUS), 0·5 m (COUTREZ, HUNAERTS, KOECKELENBERGH and probably also in the 21 cm region (MÜLLER, PRIESTER and FISCHER). It was possible to relate the 0·5 m waves unambiguously to the emissions of the CH-molecule.

While on the one hand comet Arend-Roland exhibited many interesting features, the analysis of which will contribute much to the progress of cometary research, it behaved in other respects absolutely normally. The development of its brightness was fairly typical of comets in general. The index "n" of cometary activity, which appears in the general formula for cometary magnitudes and total brightness (i.e. the absolute magnitude m_0, referred to the standard distance of 1 astron. unit), shows for this comet the same value as for most of the other comets. The maximum diameter of its nucleus was found to be about 250 km and the mass estimated from this to about 5×10^{19} g or about 10^{-8} Earth masses. These are values of the same order as those found for other comets.

Comet Mrkos appeared quite unexpectedly out of the glare of the Sun in the first days of August 1957 and by then it had already attained its greatest brightness, reaching first magnitude. Because of this it was discovered independently by many observers, but the astronomers were taken quite by surprise. Furthermore, because of its relatively unfavourable position in the night sky and of bad weather, its observation was handicapped at many places. Thus, in spite of its brilliant appearance, it furnished a much smaller amount of observational material than Comet

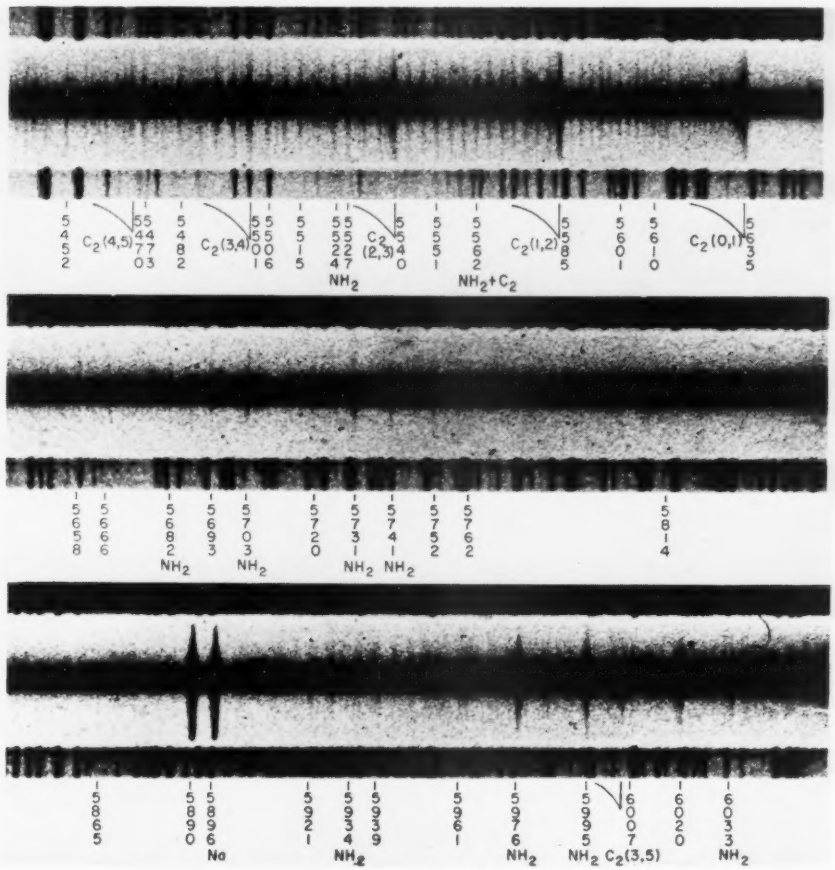


Fig. 4. Spectrum of the head of Comet Mrkos (1957d): Portion of a high-dispersion Coudé-spectrogram obtained with the Mount Palomar 200-inch Telescope. (Photo: Dr. J. L. GREENSTEIN.)

Arend-Roland. Nevertheless, it showed a surprisingly similar behaviour in certain respects. It had a strong continuous spectrum and the D-lines of sodium were prominent (Fig. 4). Observations showed the polarization to be large, but subject to considerable and sometimes abrupt fluctuations. This agrees certainly with the appearance and the behaviour of the tail of this comet. While the main tail of Comet Arend-Roland appeared relatively quiescent and uniform, and therefore offered

little opportunity for the determination of the velocities of the matter in the tail from the motion of the discrete tail-clouds, the tail of Comet Mrkos showed strong turbulence which changed its appearance from hour to hour, even from minute to minute (Figs. 5 and 6).

This is the more surprising since the development of its total brightness was by no means so strongly dependent on the distance from the Sun as in the case of Comet Arend-Roland. On the other hand, the maximum extent of its tail, about

Fig. 5. Comet Mrkos (1957d): 1957 August 21.858 U.T. Exposure time 20 minutes; Plate: Agfa Astro-Spezial; Schmidt Telescope 50/70/172 cm. Sternwarte Sonneberg. (Photo: W. Götz.)

42 million km was about the same as that of its predecessor. Objective-prism photographs showed that the different structures of the tail have to be attributed to quite different light-carriers, so that some parts of the tail are due to CO^+ ions only, while others are caused by sodium emissions. Although up to the moment only part of the detailed results concerning Comet Mrkos have been published by the observers, we may reasonably hope that the full investigation of all the data will lead to significant advances in modern cometary research.

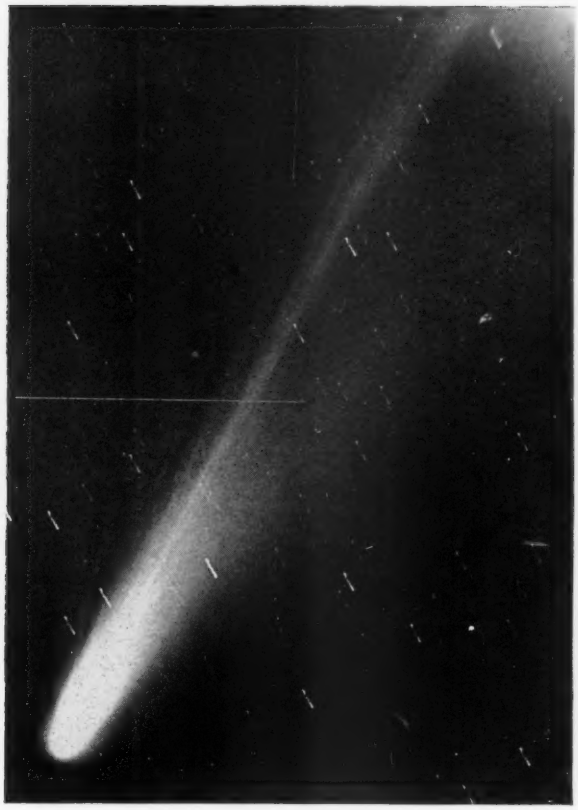


Fig. 6. Comet Mrkos (1957d): 1957 August 24.841 U.T.; Exposure time 30 minutes; Plate Agfa Astro-Spezial; Schmidt Telescope 50/70/172 cm, Sternwarte Sonneberg. (Photo: W. GÖTZ.)

III

The Anomalous Tail of Comet Arend-Roland

GUNNAR LARSSON-LEANDER
Stockholms Observatorium, Saltsjöbaden, Sweden

SUMMARY
The geometry of the tails of Comet Arend-Roland (1956h) is studied and the anomalous tail is explained as cometary matter very strongly concentrated to the orbital plane. Tabulations of the position angles of the tails and of the deviations of the tails from the prolonged radius vector are given. Finally, the origins of the anomalous tail is commented upon.

THE spectacular anomalous tail shown by Comet Arend-Roland (1956h) during the later part of April 1957 aroused great interest. It was clear very soon, however, that the main changes of the tail pattern were of a perspective nature (e.g. LARSSON-LEANDER, 1957; WHIPPLE, 1957a). The analysis given here is based on plates taken

at the Stockholm and Lund Observatories and on photographs obtained by Mr. R. FOGELQUIST, Uppsala, and by Mr. G. DARSENIUS, Gothenburg.

The significant changes in the appearance of the comet may be summarized as follows. Previous to April 25 both tails were asymmetrical in the sense that they showed weaker fan-shaped extensions on the western sides, while the eastern sides were sharply defined (Fig. 7). The position angles of the tails, anomalous tail *minus*

FIG. 7. Photograph of Comet Arend-Roland (1957 April 22, 21^h0^m—22^h0^m U.T.), showing the anomalous tail at an early stage of its development. (Photo by R. FOGELQUIST, Uppsala; 2.5 cm double-anastiginat, f = 12 cm, emulsion Ilford HP3.)

ordinary tail, differed by more than 180°. On April 25·9 the tails were symmetrical; the ordinary tail had the appearance of a parabola, and the anomalous tail appeared as a very narrow, extremely long rectilinear ray symmetrically surrounded by much weaker diffuse light. The position angles differed by almost exactly 180°. Subsequent to April 25 the tails were asymmetrical in the opposite sense, the eastern sides showed diffuse extensions, while the western sides were clearly defined. The position angles now differed by less than 180°. As the Earth passed through the orbital plane of the comet on April 25·7, it is natural to suppose that the changes are mainly due to the change of perspective, rather than to intrinsic changes of the tails.

Measurements of the position angles of the axis of the most luminous part of the ordinary tail and of the well defined rectilinear boundary of the anomalous tail are given in Table 1. The measurements are plotted in Fig. 8, where also the position angle of the Sun, at the head of the comet, is shown. The crosses give the position angle of the sharp boundary of the anomalous tail, while the appended vertical lines indicate the direction and relative size of the diffuse extension of the tail. It is evident from large-scale plates that it is this sharp boundary of the anomalous tail which is of significance for a geometrical interpretation. The estimated error of the measurements of the ordinary tail is about 2° on days when the tail was

Table 1.

Position angles of the ordinary tail and of the sharp boundary of the anomalous tail

1957	ORDINARY TAIL	ANOMALOUS TAIL	PLATE
April 22·9	2°	194°2	FOGELQUIST
25·0	12°5	194°8	,,
25·9	16°	195°1	,,
25·9	15°8	195°0	Stockholm
26·9	20°4	194°0	,,
27·9	25°	192°5	,,
28·9	27°8	192°	,,
29·9	32°	190°	Lund
30·9	35°		DARSENIUS
May 1·9	37°9		Stockholm
2·9	39°5		,,

relatively wide (April 22·9, May 2·9) or distorted (April 27·9), but the accuracy is greater on other days. The error of the measures of the boundary of the anomalous tail is generally below 1°.

From the measured position angles the deviations, φ_1 and φ_2, of the axis of the most luminous part of the ordinary tail and of the sharp boundary of the anomalous tail, respectively, from the prolonged radius vector were calculated by means of Bessel's formulae (JAEGERMANN, 1903, p. 83). It was assumed in these calculations that the orbital plane was a plane of symmetry, as is indicated by the symmetrical tail pattern observed on April 25·9. Due to the proximity of the Earth to the orbital plane of the comet the values of φ_1 and φ_2 are very sensitive to the errors of measurements; in fact the errors of the position angles are multiplied by a factor of about 5 when transformed to the orbital plane. In order to minimize this effect, the results given in Table 2 are based not on the position angles actually measured but on values read from smooth curves drawn through the dots and crosses in Fig. 8.

The geometry of the tails is illustrated in Fig. 9. The ordinary tail is schematically depicted as a cone, but in reality it was slightly curved. Due to the curvature the determination of the "axis" of the tail is somewhat ambiguous. As the position angles were determined by considering points at some distance from the nucleus the values φ_1 given in Table 2 are larger than the values valid for the immediate neighbourhood of the nucleus. A more correct representation of the ordinary tail than that of Fig. 9 would therefore show curved structures having the same general direction, but being more nearly tangential to the prolonged radius vector close to the nucleus. On the side away from the prolonged radius vector the ordinary tail

had an extension of matter more strongly concentrated to the plane, but probably the main part of the tail was also somewhat elongated in the orbital plane. The anomalous tail, on the other hand, was formed of matter *very* strongly concentrated to the orbital plane. From plates taken with the Stockholm Observatory 60 cm refractor on April 25·9, when this tail was seen edge-on, its width, not considering the diffuse envelope was found to be about 30″, which corresponds to a thickness of about 13,000

Table 2

The deviations, γ_1 and γ_2, of the axis of the ordinary tail and of the sharp boundary of the anomalous tail from the prolonged radius vector.

1957	γ_1	γ_2
April 22·9	21°	94°
27·9	45°	98°
29·9	51°	108°
May 2·9	44°	
6·9	10°	

km. On the side towards the Sun the anomalous tail was sharply bounded but was more diffuse in the opposite direction. Its extension in the orbital plane towards the ordinary tail was considerable. Thus, on April 22·9 the angle of aperture of the tail sector was about 18°, which corresponds to about 24° in the orbital plane. Some later photographs, moreover, seem to indicate that the whole region between

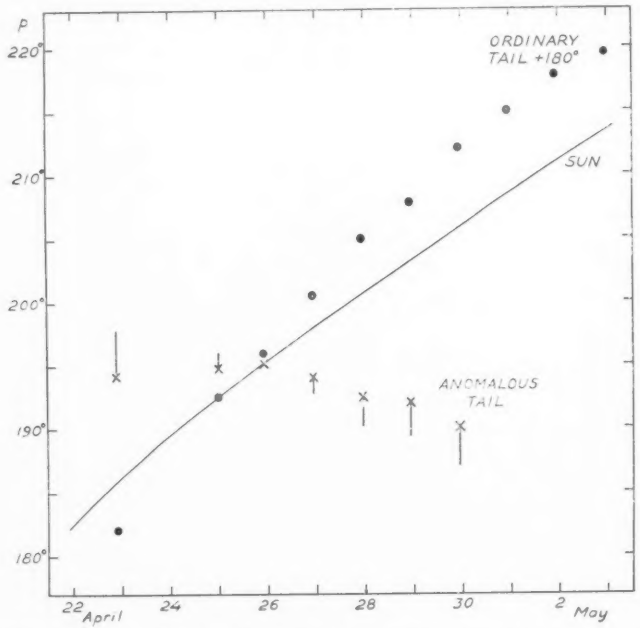


FIG. 8. Measured position angles of the axis of the ordinary tail and of the sharp boundary of the anomalous tail (crosses). The vertical lines appended to the crosses indicate the direction of the diffuse extension of the anomalous tail. The calculated position angle of the Sun, from the head of the comet, is shown by the continuous curve.

the two tails was filled by diffuse light. On the plate obtained by FOGELQUIST on April 25·9 the rectilinear ray of the anomalous tail may be traced for at least 13°, and it seems probable that this limit was set mainly by the increasing opacity of the atmosphere close to the horizon. From this observation it follows that the linear extension of the anomalous tail along its sharp boundary was at least $25 \cdot 10^6$ km.

As pointed out by several authors (NEEDHAM, BEER and HO, 1957; VORONTSOV-VELYAMINOV, 1957; WHIPPLE, 1957b) a number of comets have shown anomalous tails of about the same character as Comet Arend-Roland. No previous observations are known, however, of a comet exhibiting a spike-like anomalous tail of a length comparable with the present one.

Among the more thoroughly observed previous objects Comet 1823 is of special interest for a comparison with Comet Arend-Roland. The transit of the Earth through the orbital plane of Comet 1823 occurred on 1824 January 22, and the anomalous tail was observed from this day until January 31. When brightest, the anomalous tail was quite comparable to the ordinary one, both with regard to intensity and length. From the descriptions and drawings of the appearance of the tail, given by VON BIELA (1824), by HARDING (1824), and by OLBERS (1824), it is obvious that

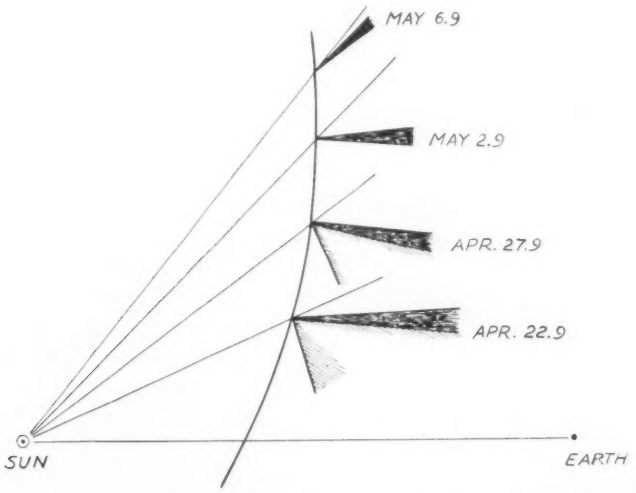


FIG. 9. The geometry of the tails. Shaded areas represent matter strongly concentrated to the orbital plane. The position of the Earth is that of April 25·7, the day of passage through the orbital plane of the comet.

the concentration of cometary debris towards the orbital plane was in this case less marked than in that of Comet Arend-Roland. BREDICHIN (see JAEGERMANN, 1903, p. 457) concluded from his study that the direction of the anomalous tail almost coincided with the tangent to the path of the comet. It thus followed the radius vector; the tail was a pseudo-anomalous tail according to BREDICHIN's terminology. (In this terminology the true anomalous tails are directed more or less towards the Sun but are preceding the radius vector.) As the matter forming the anomalous tail in this case moved in the same orbit as the nucleus, the whole assembly was regarded as a train of meteorites, the largest of which was the nucleus of the comet.

No forces other than the gravitational attraction were necessary to explain the phenomenon.

In the case of Comet Arend-Roland the pseudo-anomalous tail strongly deviated from the radius vector and also from the tangent to the orbital hyperbola of the nucleus. The orbits of the individual particles forming the tail are thus hyperbolas with eccentricities larger than that of the cometary orbit. This fact and the observed large extension of the tail in the orbital plane, as well as its sharp boundary towards the sunward side, indicate that the matter forming the anomalous tail was subject to a repulsive force of solar origin, which, however, was much weaker than the forces acting upon the particles of the ordinary tail. The large concentration of particles to the orbital plane shows that the repulsive force was acting in this plane. This seems to preclude the idea that corpuscular radiation could be responsible, as the paths of emitted solar particles would not be radial at the distance of the comet. Radiation pressure on meteoric dust may be a possible explanation or, at least, part of the explanation. Another additional mechanism, jet action due to solar heating of the sunward side of small meteoric particles covered with ices, has been suggested by WHIPPLE (1957 a, b).

From the large-scale photographs it seems evident that the main part of the matter forming the anomalous tail was ejected from a region very close to the nucleus. If this ejection occurred in all directions, as seems plausible, the velocity of most of the ejected particles must be very small compared with the comet's orbital velocity. Otherwise the large concentration of matter to the orbital plane could not have been maintained. It should be emphasized, however, that the diffuse halo around the spiked anomalous tail on April 25·9 shows that particles were present also at considerable distances on both sides of the plane.

The writer is indebted to Professor C. SCHALÉN, Director of Lund Observatory, to Mr. G. DARSENIUS and Mr. R. FOGELQUIST, as well as to colleagues at this observatory, especially Mrs. K. LODÉN and Mr. L. O. LODÉN, for placing observational material at his disposal. He is also grateful to Mr. P. O. LINDBLAD for his calculations of the directions in space of the tails of Comet Arend-Roland at an early stage of the present investigation.

REFERENCES

BIELA, W. v.	1824	*Astron. Nachrichten*, **3**, 27.
HARDING, K. L.	1824	*Astron. Jahrbuch*, 1827, p. 131, Berlin.
JAEGERMANN, R.	1903	*Prof. Dr. Th. Bredichin's mechanische Untersuchungen über Cometenformen*, St. Petersburg.
LARSSON-LEANDER, G.	1957	*U.A.I. Circ.* No. 1599.
NEEDHAM, J., BEER, A. and HO, PING-YÜ	1957	*Observatory* **77**, 137.
OLBERS, W.	1824	*Astron. Nachrichten*, **3**, 5.
VORONTSOV-VELYAMINOV, B. A.	1957	*Astron. Circ. U.S.S.R. No. 180*, p. 6.
WHIPPLE, F. L.	1957 a	*Nature (London)*, **179**, 1240.
	1957 b	*Sky and Telescope*, **16**, 426.

IV

Filter Photography of the 1957 Comets

NEVILLE J. WOOLF
The Astronomy Department, Manchester University

AFTER many lean years 1957 provided a feast of two bright comets that prodigally came to perihelion within four months of one another. Nevertheless they are dynamically unrelated. Because of the relative Earth–Sun–Comet positions MRKOS (1957d) was only discovered at perihelion whilst AREND-ROLAND (1956h) was found six months before its perihelion passage. The apparent tail structures also were very different. Comet Arend-Roland seen from near the plane of its orbit had concentrations of matter in this plane accentuated, whilst for Comet Mrkos the line of sight was nearly perpendicular to the plane of its orbit so that it showed separate gas and dust tails, but concealed any "beard" or "aiguillon" such as the Arend-Roland comet showed.

Shortly after the spectacular beard of the Arend-Roland comet had disappeared it was explained geometrically by LARSSON-LEANDER (1957) and WHIPPLE (1957) as a concentration of debris in the orbital plane. Physically however, the phenomenon is still obscure. The beard showed a sharp sunward edge and a diffuse trailing edge similar to the proper tail. There was no sign of any structure to suggest the presence of gas. If the sharp sunward edge was the locus of miniature nuclei emitting the dust which was the main part of the beard, it is not easy to see how these nuclei were separated from the main nucleus and why they were in line in the comet's orbital plane.

Comet Mrkos' tail showed a well marked division into a nearly straight gaseous tail and a broad curved dust one. A feature of the dust tail warranting more attention was its peculiar streaked appearance on 13 and 14 August 1957. The gas tail emitted light of the usual CO^+ bands and also the sodium D lines. Filter photography by BIGAY, DOAN and DUFAY (1957) separated these two emissions and showed that whereas the ionized carbon monoxide streamers were of the usual wavy appearance, the neutral sodium emission was concentrated in bright straight streaks radiating from the nucleus a distance of $5^\circ 5$ into the tail. The sodium emission at large heliocentric distances was extraordinarily strong in both comets, yet it seems improbable that the abundance of sodium should be anomalous in both.

Because the continuous spectra of the comets were strong it may be possible to determine the nature of the "dust" that gave rise to them. The spectral distribution of this scattered sunlight has been studied photoelectrically and it has been shown to be reddened. The particles were therefore the same size as, or somewhat larger than the wavelength of light. The ratio of solar gravitational pull to radiation pressure on the particles can be determined from the position angle of the tails; also the polarization of the continuum at various scattering angles has been measured. Together these three measures should elucidate the nature of the dust. The new polarization measures by BLACKWELL and WILLSTROP (1957) and MARTEL (1958) are important as not only has there been little cometary emission to dilute the

polarization by an uncertain amount, but also observers have used colour filters to prevent this. Interpretation of the results will be difficult since, the particle-size being of the order of the wavelength of light, measures on macroscopic objects give invalid comparisons.

Many observers photographed the comet through filters to isolate the emission of the separate bands, but this technique is not a panacea. The origin of the CO^+ tail-streamers is not known and requires the study of these streamers near the cometary nucleus. Normally the brightness of the molecular emission bands prevents this, and therefore A. N. Argue of the Cambridge Observatories and the writer tried to isolate the CO^+ emission of the Arend-Roland comet using a 50 Å half-width interference filter. Together the strength of the continuous spectrum and the weakness of the CO^+ emission made this project unsuccessful. One might suppose that narrow band filters could always improve the contrast of the streamers, but this is not the case. The brighter CO^+ bands come in pairs about 22 Å apart, the pairs being separated by about 200 Å. As one contemplates the use of filters of decreasing bandwidth, one sees that CO^+ bands will be rejected nearly as rapidly as the continuum. Because the bands are in pairs, even the narrowest bandwidth interference filters (about 15 Å) will not help. When the continuum, rather than the band emission of the head, is the enemy, it appears that interference filters can only improve the streamer contrast by a factor of about 3. Short exposures are needed as one is interested in the changes of the streamer structure, but the rejection of much of the CO^+ light by the filter prevents this. A 30 min. exposure through the 50 Å filter with a telescope of effective focal ratio f/4 showed faint streamers that could only be traced a little nearer the nucleus than on a $7\frac{1}{2}$ min. unfiltered exposure. With gaseous comets the technique should be feasible; for comets with much dust the optical problem seems intractable. Studies of the head emissions by others will probably have suffered from the same difficulty, though less acutely.

In view of the coincidence of Comet Mrkos with the International Geophysical Year it is hopeful that it will help to elucidate the correlation between cometary brightness and solar activity. Work by Thiessen (1957, 1958) on Comet Arend-Roland showed a correlation between decrease of cometary brightness and an increase of sunspot numbers, whilst previous work by Beyer (1952) on other comets suggested a correlation with increasing brightness, indicating a higher probability of discovery of a comet near sunspot maxima.

REFERENCES

Beyer, M.	1952	*Mem. Soc. Roy. Sci. Liège*, **13**, 236.
Bigay, J. H., Doan, N. H. and Dufay, M.	1957	*C.R. Acad. Sci., Paris*, **245**, 921.
Blackwell, D. E. and Willstrop, R. V.	1957	*M.N.*, **117**, 590.
Larsson-Leander, G.	1957	*Observatory*, **77**, 132.
Martel, M.	1958	*C.R. Acad. Sci. Paris*, **246**, 2347.
Thiessen, G.	1957	*Z. Astrophys*, **43**, 260.
	1958	*Z. Astrophys*, **44**, 169.
Whipple F.	1957	*Nature (London)*, **179**, 1240.

PART II

STELLAR ASTRONOMY
PHOTOMETRY
SPECTROSCOPY
STELLAR EVOLUTION
GALAXIES
COSMOLOGY

VOL
3
196

The First 70,000

JOHN B. IRWIN

Goethe Link Observatory, Indiana University, Bloomington, Indiana, U.S.A.

SUMMARY

A brief examination and discussion is given for the nine types of basic observational data which may be quickly obtained with telescopes of moderate size for stars brighter than visual magnitude 8·5. It is concluded that for these 70,000 stars photoelectric multicolour data and much more accurate proper motions are most urgently required. Furthermore, existing photoelectric techniques should be applied to the determination of their precise spectral and luminosity classes. The development of a photoelectric technique for the quick determination of radial velocities offers a special challenge and seems attainable. The key rôle of the photoelectric cell in most of these types of observations strongly suggests the need for a new and specialized type of observatory; such an observatory is especially needed in the southern hemisphere.

1. INTRODUCTION

THAT astronomy is something of an observational science is no news to anyone who has contemplated the hundreds, if not thousands, of thick volumes that may be found in any good astronomical library which contain the results of observations of one kind or another. In view not only of the continually changing directions and emphasis of astronomical and astrophysical research, but also of the development and manufacture of powerful new instrumental accessories, it would seem to be profitable from time to time to examine, however briefly, both the needs and possibilities in regard to the mass production of different kinds of observational data of the highest possible quality. This paper is an attempt at such an examination and is concerned with and limited to those types of observations that may be made with adequate precision in a few minutes per star with telescopes 36 inches in diameter or less for stars brighter than visual magnitude 8·5. There are approximately 70,000 such stars and this somewhat arbitrarily selected group is chosen only to fix ideas and as a basis for discussion. Telescopes larger than 36 inches in diameter are so expensive and so unwieldy on the one hand and of such great usefulness in attacking critical problems concerned with fainter objects on the other, that their extensive use in the routine production of data might well be seriously questioned.

One of the most active fields of investigation in modern stellar astronomy is that of stellar evolution, and substantial and exciting progress is at long last being made in this difficult subject. This progress has come about through our increased knowledge of stellar energy generation derived from studies in nuclear physics and has been further accelerated especially by studies of the colour-magnitude diagrams of both globular and galactic clusters. The globular cluster studies, stimulated by BAADE's stellar population concepts, were made possible by the advent of the efficient multiplier phototube. These studies are almost exclusively a field of investigation for the very large reflector and it is to be hoped that the fundamentally important work of the Palomar observers may be checked and supplemented by

observations with our all too few other large reflectors; furthermore, extension of this work to the southern hemisphere would be especially desirable, not only because two-thirds of our Galaxy's globular clusters are south of declination $-20°$ but also because of the presence there of both the Magellanic Cloud globulars and such important nearby clusters as 47 Tucanae and Omega Centauri. Photoelectric studies of galactic clusters are perhaps even more important in advancing our knowledge of stellar evolution and here the small reflector can be and has been used most effectively. An examination of SHAPLEY's (1933) list shows 44 galactic clusters south of $-20°$ whose brightest stars are brighter than tenth photographic magnitude and whose estimated distances are within 1,200 parsecs; these clusters have scarcely been touched.

Perhaps even more important than either type of cluster are the associations, because if an association can be identified and its expansion measured, then a most important piece of information, namely the lifetime of its component stars, is obtained. Given enough such information derived from many such associations one may hope to answer eventually the following important question in considerable detail, namely: How do the various observable properties of a star such as its temperature, size, luminosity and chemical composition change during its lifetime? Or, from the point of view of a theoretical astrophysicist calculating evolutionary models of stars, knowledge of the lifetime of a star is as necessary as knowledge of M, L and R in order to guide the calculations to results that have physical meaning.

An inspection of the MORGAN, WHITFORD and CODE (1953) list of 27 associations of high luminosity stars discloses the kinds of data we need to know in order to study such groups. In order to obtain trustworthy distances one needs reliable luminosities combined with apparent magnitudes corrected for absorption; this correction is accomplished by means of colour excesses derived from multicolour observations alone or in conjunction with spectral types. The age of the association can be determined in principle either by means of the differential proper motions or the differential radial velocities. Most of these 27 associations are distant 1,000 parsecs or more, yet they contain many naked eye stars whose proper motions are usually not as yet accurately enough known for expansion ages to be determined. Fig. 1 illustrates the proper motion data as given by DELHAYE and BLAAUW (1953) for both right ascension and declination for the nearest of these associations. Proper motions ten times more accurate should make this expansion phenomenon crystal clear and might even answer other important questions such as whether the stars were all born at the same time and in the same place. The study of such associations is still in its infancy and most of the necessary data—even for the brighter stars—as to proper motions, colours, magnitudes, spectral types, luminosities, distances and radial velocities have not as yet been published or even obtained with the precision now possible with modern techniques.

A major question concerning these associations is the identity of the member stars, and here our basic data are too imprecise and need substantial strengthening. Modern data on all stars brighter than a given magnitude limit would locate many new members and almost certainly new and new types of associations. At one time (RUSSELL, 1952) it was enthusiastically "anticipated that the general mass of stars of known proper motions would, after sufficient study, be divisible into an assemblage of moving clusters, having relatively few 'mavericks' which failed to belong. We now know that this hope was vain". Given accurate enough space motions for

70,000 stars however, we might still hope to be able to pinpoint the birthplace and date of many thousands of the more luminous of this group and pick out those stars needing high dispersion spectroscopic attention with the great reflectors. To put it another way, we don't know which stars we are interested in until after the best possible basic observational data has been obtained and analyzed.

The above considerations suggest eight kinds of basic data needed, to which we will add polarization. The list to be discussed is then:

1. Position.
2. Proper Motion.
3. Magnitude.
4. Multicolours.
5. Spectral Type.
6. Luminosity Class.
7. Parallax.
8. Radial Velocity.
9. Polarization.

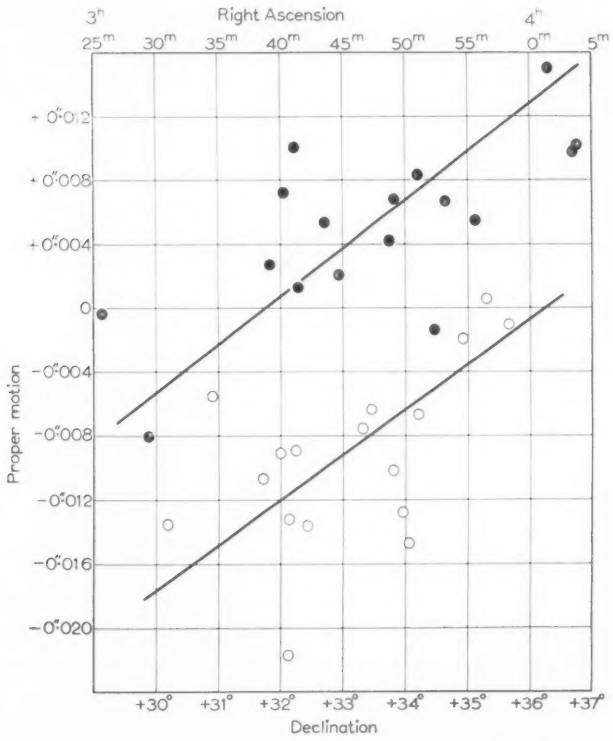


FIG. 1. Proper motions in right ascension (filled circles) and declination (open circles) for members of the ζ Persei association. The slopes of the lines correspond to an expansion age of $1{\cdot}3 \times 10^6$ years.

There are positions, proper motions, magnitudes and spectral types of sorts for essentially all the stars brighter than visual magnitude 8·5; radial velocities for about 13,000 stars and multi-colours, luminosity classes, parallaxes and polarizations for only a relatively small percentage of our 70,000 stars. However, if we consider standards of accuracy which are presently obtainable with ease with moderate-size telescopes then it turns out that over 99 per cent of the stars need attention in most of

these divisions. It should be pointed out that this discussion does not include such things as stellar magnetism, rotation, duplicity, variability, emission lines, chemical composition, spectrophotometry, radio astronomy, sky-mapping, or observations of planets, nebulae, and galaxies.

2. Positions and Proper Motions

The Boss General Catalogue of 33,342 stars, published in 1937, is the present day source of most of the accurate proper motions (for stars brighter than visual magnitude 8·5) but not of their positions. Although the assembling of this catalogue, hereafter called the GC, was a magnificent achievement in bringing together and making full use of millions of meridian observations, it was immediately recognized that the catalogue had some obvious weaknesses. Perhaps the most serious of these concerns the positional accuracy of the fainter stars, a condition worse today than 20 years ago because of the increased length of time of extrapolation from the mean epoch of observation which is about 1900. A fundamental system is defined by the constant of precession and the positions and proper motions of a network of stars. Over two-thirds of the GC stars now have probable errors greater than $\pm 0''.40$ and the system is much too fuzzily defined for stars fainter than sixth magnitude. The average positional errors for 1960, increased by 15 per cent in conformance with the Yale Observatory practice, are plotted in Fig. 2 as a function of apparent magnitude. The variation of error from star to star within any narrow magnitude interval is substantial. The poorer quality of the southern hemisphere positions is apparent, especially for the stars south of $-30°$. The sharp increase in errors between fifth and sixth magnitude is an outstanding feature. The published probable errors for both right ascension and declination have been combined in this graph; in general the right ascension errors are approximately 13 per cent larger than the declination errors. If a position with a probable error of $\pm 0''.05$ were to be obtained in 1960 for the stars in the catalogue, the resulting errors for 1970 and the year 2000 are shown, averaged for the whole sky. For the year 2000 the improvement is a factor of 4 as compared to the present time and the probable error of $\pm 0''.12$ is comparable with that of the Yale Zone Catalogues at their respective epochs. The present day errors of the Yale positions vary from $\pm 0''.16$ to $\pm 0''.34$ depending on the declination zone and refer, of course, to many more stars than are in the GC. The GC is essentially complete to visual magnitude 7·0 so that it might be expected that the errors in position and motion for seventh and eighth magnitude stars not in this catalogue would be somewhat larger than shown in Fig. 2. This may or may not be so, depending again on the declination zone. For example, in the Yale Zones from $+30°$ to $-30°$ the annual proper motion probable errors range from $\pm 0''.004$ to greater than $\pm 0''.013$.

There are two obvious methods by which the GC errors may be substantially reduced; one is by further observation, and this is being done, especially in the AGK3 programme (Heckmann, 1954). The second method is by properly absorbing the great quantities of stellar positions that have been published since the early 1930's, and this is *not* being done. Without going into any detailed argument at this point, it would seem to be obvious that both approaches are necessary and desirable; and further, an improved GC should include stars to a fainter limit and tentatively, this limit should be visual magnitude 8·5.

The reasons we need such an improved and enlarged GC are many and are substantial. One such reason is the desire for more accurate proper motions for our 70,000 stars arising from our need to know their tangential velocities more accurately. There are two kinds of problems here; on the one hand we need percentage accuracies

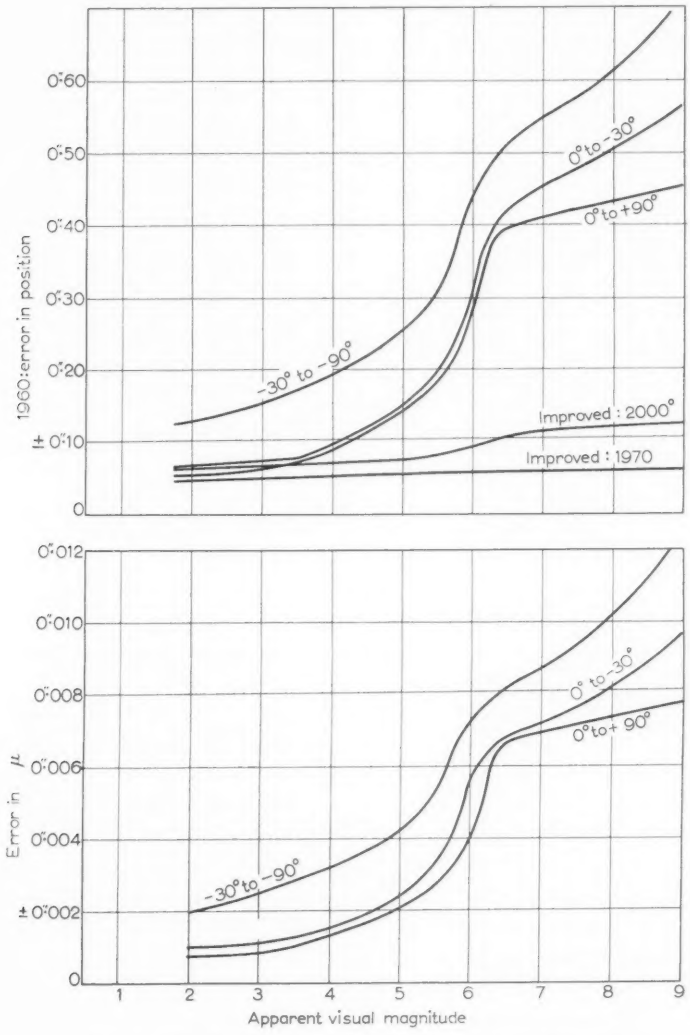


FIG. 2. Probable errors of the 1960 GC positions (above); probable errors of the GC proper motions (below).

of the order of 5 or 10 per cent in any discussion of moving cluster or expanding association characteristics, while in discussions of statistical parallaxes the errors that can be tolerated might be 10 times greater. In the former case, systematic errors of the fundamental catalogues might not be important—with the exception of magnitude errors which *are* important; in any particular problem the stars under consideration would usually be in a relatively small region of the sky. Errors in the parallax

might or might not be important here. In the case of the determination of statistical parallaxes, systematic errors would be more serious. The average tangential velocity for the stars of brightest apparent magnitude is about 22 km/sec. The nearest early B-stars, Cepheids and supergiants would average less than this while the O-stars would average somewhat higher. We will here accept a tangential velocity as accurate enough for statistical purposes if the tangential velocity probable error due to the proper motion error alone, is equal to or less than ± 15 km/sec; a "precise" tangential velocity is here defined as one with one-tenth this error or $\pm 1{\cdot}5$ km/sec. At any magnitude level the error requirements are such as to be most stringent for the stars of highest luminosity; furthermore, the requirements become increasingly more stringent as we go to fainter apparent magnitudes. It is convenient to discuss three different absolute luminosity levels, namely: $M_v = 0{\cdot}0$, $-3{\cdot}5$ and $-7{\cdot}0$, corresponding approximately to the absolute magnitudes of B9V or K0III stars, B1V or an average classical Cepheid, and Ia supergiants, respectively. Distances have been calculated at appropriate apparent magnitudes for stars of these three absolute luminosities; a visual absorption of 1 mag/kpc has been assumed. The proper motion probable errors corresponding to both the "statistical" and "precise" cases were then calculated from the familiar equation, $\mathbf{v} = pT/4{\cdot}74$, with T equal to ± 15 km/sec and $\pm 1{\cdot}5$ km/sec respectively.

The resulting six error curves are shown in Fig. 3 along with the GC proper motion error curve for the whole sky. The six intersections of the former with the latter give "precise" and "statistical" apparent magnitude limits. For example, GC stars brighter than visual magnitude 8·2 will, on the average, have errors in μ small enough to be useful for statistical purposes provided M_v is equal to or fainter than 0·0; on the other hand Ia supergiants will only have precise proper motions for those very few stars brighter than first magnitude. The intersection point of the $M_v = -3{\cdot}5$, $T = \pm 15$ curve at visual magnitude 5·9 is in reasonable agreement with the experience of Blaauw and Morgan (1954) who found that only 18 of the brighter Cepheids were statistically useful in fixing the zero-point of the period-luminosity law; most of the numerous cepheids fainter than magnitude 6·5 had proper motions so inaccurate—even with improved data—that they were useless for statistical purposes.

The lower curve labelled "Improved GC Motions" is the result again of combining GC data with hypothetical positions obtained in 1960 with probable errors of $\pm 0''.05$. The improvement in accuracy is more than a factor of 3 for stars fainter than sixth magnitude and somewhat less than this factor for brighter stars. The improvements in the magnitude limits are listed in Table 1.

Table 1

	STATISTICAL ($T = \pm 15$ km/sec)	PRECISE ($T = \pm 1{\cdot}5$ km/sec)
$M_v = 0{\cdot}0$	$8^m.2$ to $>10^m.0$	$5^m.2$ to $6^m.3$
$M_v = -3{\cdot}5$	$5^m.9$ to $>8^m.5$	$3^m.3$ to $4^m.3$
$M_v = -7{\cdot}0$	$4^m.8$ to $>6^m.1$	$1^m.0$ to $1^m.7$

An improvement of one magnitude would correspond to an increase in numbers of stars by a factor of about 2·8 on the average and 2 magnitudes would correspond

to a factor of about 8; it would not be this much for the highly luminous stars. Thus, we might reasonably expect a very substantial and useful extension of our knowledge of stellar motions to many more stars, provided the GC proper motions were to be improved by 1960 positions in the manner previously indicated. It should be emphasized however that this discussion does not apply to high-velocity stars nor to stars brighter than $8^{m}5$ which are not in the GC, and further, has ignored the substantial differences in northern and southern accuracies.

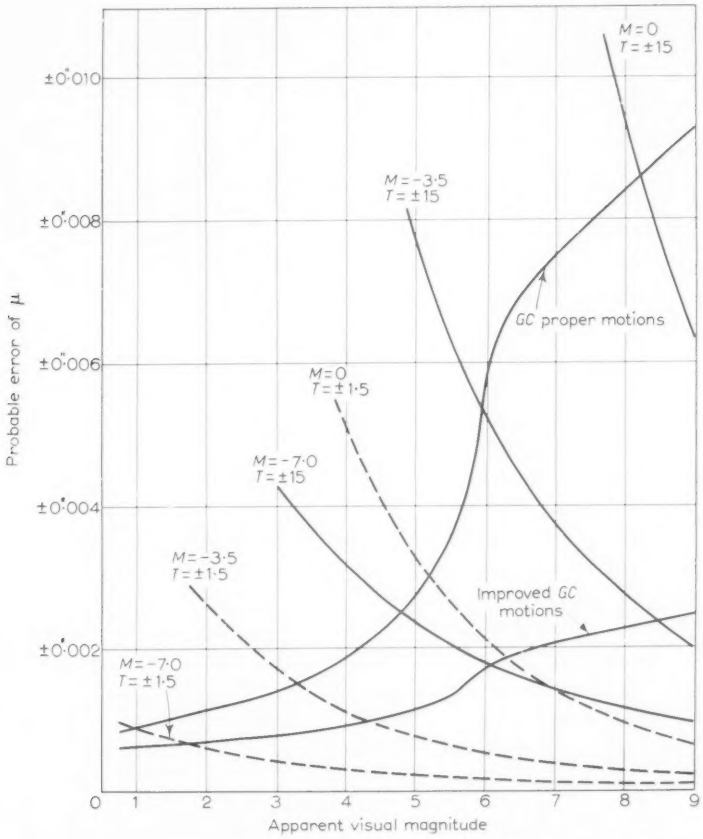


FIG. 3. Annual proper motion errors of the GC. The solid lines ($T = \pm 15$ km/sec) represent "statistical" accuracies; the dashed lines ($T = \pm 1{\cdot}5$ km/sec) represent "precise" accuracies. The intersections of these six curves with the "GC proper motion error" curve give the magnitude limit to which these accuracies hold for $M_v = 0{\cdot}0$, $-3{\cdot}5$ and $-7{\cdot}0$. The "Improved GC motions" represent the effects of adding an observation at 1960 with a probable error of $\pm 0''.05$.

Although it is difficult to guess what improvements in accuracy might be expected if the GC were to be brought up to date, there is no question but that such should be done as soon as possible. No observational catalogues are included in the GC that were published later than the early 1930's and excellent positions observed 20 years before this deadline were, of necessity, omitted. The number of observational catalogues that could be included in a revised GC would number close to 100; most of

these would be catalogues of modern precision and some of them are very extensive. The AGK2 volumes, the Yale Zone Catalogues, and the numerous and excellent Cape Photographic Catalogues could be included, for example. The FK4 is about to be published and this would provide a modern, definitive framework upon which to build. Putting aside for the moment the question of AGK3 positions and other possible new positions, a discussion of observations already made would give the best possible positions at a mean epoch of about 1925 *that could be obtained in no other way*. A decisive factor in initiating such a large project would be the presence at some of the large national observatories of modern data-processing machines. Such computing aids could handle and process and analyse these copious data effectively and quickly and, in fact, this is exactly the type of problem for which such machines are built.

An extension of the magnitude limit to 8·5 would make any revision of the GC a much more valuable one. Such extension would make better and more complete use of existing star catalogues—such as the AG catalogues for example—and would enable better systematic errors to be derived for these catalogues, especially for the fainter stars. It would extend our knowledge of definitive proper motions to more than twice as large a group of stars, placing them on a homogeneous system with substantially reduced systematic errors. The resulting network of 1·7 stars per square degree on the average would be especially useful, with proper supplementation at high galactic latitudes, in providing the reference stars necessary for extension of this system, by photographic interpolation to fainter stars. From the point of view of this paper it might, for example, lower the "Improved GC Motions" curve in Fig. 3 by 30 to 40 per cent, thus making available both "statistical" and "precise" motions for a considerably larger number of stars.

The primary instrument used in the determinations of absolute positions and proper motions is the transit circle. Modern requirements of accuracy are so high that it would seem profitable to subject this instrument to the closest scrutiny in the hope that it can be refined or modified in the direction of much greater accuracy. In the GC, an observation of unit weight is one with a probable error of $\pm 0''.30$ and none of the more than 200 catalogues listed in Volume 1 of the GC achieve weights of greater than 1 for single observations. This precision was first reached in the Pulkovo 1845 vertical circle observations for declination and some 45 years later by a number of observatories in the right ascension co-ordinate. There have been cases of some recent, modest improvements; the right ascension observations for the 1936–41 6-inch Washington transit circle yield probable errors for single observations of $\pm 0''.20$ over most of the sky, even at zenith distances of 50° (Watts and Adams, 1949). This improvement was undoubtedly due to the use of a superior motor-driven micrometer combined with a fairly long (6 observations spaced at 4-second intervals) time of observation on each star at its meridan passage. The declination errors average $\pm 0''.30$ or a little worse and here we have an unusual situation in that the difficult moving co-ordinate is measured more precisely than the stationary co-ordinate. In the case of the 1932–34 zone (−10° to −20°) observations of the 9-inch Washington transit circle (Morgan and Lyons, 1937) the situation is reversed, $\pm 0''.34$ in right ascension (an impersonal micrometer was not used) and $\pm 0''.23$ in declination. The latter remarkably low figure, especially for zenith distances of 50° to 60°, was confirmed by the Yale Zone Catalogue discussion; the reasons for it should be of thoughtful interest to meridian observers. The value of the probable

error of a single observation at these zenith distances was $\pm 0''.30$ in a later programme with the same instrument (MORGAN and SCOTT, 1948).

All these quoted probable errors are larger than can be obtained from one photographic observation ($\pm 0''.15$) and even larger as compared to photographic zenith tube results which are of the order of $\pm 0''.10$. The PZT has certain natural advantages—zenith observations only and an absence of azimuth, level, collimation and refraction errors. It is possible to observe stars widely separated in the sky—in right ascension—to this high accuracy in both co-ordinates. The question naturally arises as to why this is not possible with transit circles for stars separated widely in declination—or at least in 10° zones of declination—in either co-ordinate, *even in the zenith*. The answer to this question is of basic importance if we would hope to achieve the precision in meridian work that we badly need as soon as possible. The writer's guess is that a star should be observed in its meridian passage for as long as a minute (see, for example, LAND, 1954 and SCHLESINGER, 1916) and, more to the point, a different type of observational technique should be used. The PZT uses photographic registration but it would appear to be better to use continuous or nearly continuous photoelectric guiding if such is possible. WEITBRECHT's (1957a) photoelectric guider goes to 9th magnitude with an $18\frac{1}{2}$-inch telescope, comparable to $6^m.7$ with a 7-inch. A gain of a factor of 2 can be had if the star's light is split two ways rather than four and undoubtedly further refinements and experimentation—perhaps a better phototube selection—could reach $8^m.0$ or $8^m.5$ with a 7-inch. REIZ (see 1955 *I.A.U. Transactions*, p. 113) has successfully tested in the laboratory such a device for a transit circle. A move to clearer and more transparent skies would be helpful in getting that last half magnitude and is desirable in any case and possible in some cases such as the U.S. Naval Observatory instruments, for example. There are many problems both of construction and of observational routine that need to be solved for photoelectric micrometers. Once a successful solution is obtained it is to be hoped that its universal adoption will not emulate the incredibly slow acceptance of the motor-drive impersonal micrometer.

One of the difficulties encountered is that transit circles are highly sensitive to both mechanical and temperature effects. An example of temperature sensitivity is the now-retired 9-inch Washington transit circle (MORGAN, 1906; MORGAN and SCOTT, 1948) which nevertheless produced observations of modern precision as previously mentioned. All transit circle buildings go through the same cycle in that the east, south and west walls (in the northern hemisphere) are successively heated throughout the day—and also the roof—and this has marked daily and seasonal effects on the so-called instrumental constants. The walls and roof could undoubtedly be better protected and kept "shadowed" to reduce such systematic effects. Fundamental programmes, have in the past, included observations of the Sun and planets; the Sun is surely one of the poorest objects to observe for position that we have and the hope is that observations of selected asteroids may avoid the necessity of making solar observations. The Sun is the last object in the sky that one would want to have shine on a delicate instrument! Observers who may well be dubious that transit circle observation can be improved by factors of 2 and 3 by a change of observing techniques might recall the real "break-through" in the determination of lunar positions with the MARKOWITZ (1954) dual-rate moon camera with which the probable error of a single observation has been reduced by a factor of over 3, and with which the moon may be observed over a greater orbital arc. It may well be that planetary

observations could be similarly abandoned by transit circles and observed much more accurately—as they should be—only by long focal-length refractors. For transit circles we are interested in quantities of the order of one hundredth of a second of arc and this corresponds to only 0·02 microns on the rim of a circle of 32 inches diameter. The builder of such an instrument must therefore scrutinize every detail of the design and construction with such limits in mind. Instrument makers today have at their disposal materials, machines and testing techniques that were unknown fifty years ago and much better and more stable instruments should be possible. However, as Sir DAVID GILL wrote in a letter to KAPTEYN (FORBES, 1916): "But however perfect an instrument may be (and it is the astronomer's business to see that it is perfect), it is the astronomer's further business to look upon it with complete and utter mistrust". Meridian astronomers should not be content with the performance of their instruments until such time as they measure up to the performance of the PZT, at least at declinations near the zenith. It is certain that the troubles are in the observing methods and the instruments and are not in the "seeing" except possibly for extremely large zenith distances.

The GC accuracies are poor in the southern hemisphere, but would have been much poorer yet if it had not been for the Dudley Observatory's 1909 expedition to San Luis, Argentina, where 60,000 meridian observations were made in one year. It is well known that much more astrometric effort must be put into the southern hemisphere work in order to reduce both accidental and systematic errors in our fundamental system and in extension of that system to fainter objects. Even with the best of intentions and with an established working organization it takes a very long time to get an entirely new transit circle into full-time operation. For example, it is 13 years since money was appropriated for the new 7-inch transit circle at Washington and 27 years since the Greenwich 7-inch Cooke RTC was started. Neither instrument has as yet produced published observations and may not do so for some years to come. The point is that if we wish well-investigated modern transit circles and experienced observers in the southern hemisphere, the quickest and most direct way of supplying this need would be by a number of expeditions similar to the San Luis expedition. It may be that such expeditions could be planned now at some of the large national observatories and action taken as soon as the AGK3 programme observations have been completed.

The objectives of the AGK3 programme have been well stated by HECKMANN (1954). Briefly, simultaneous transit circle and photographic coverage of the northern sky to -5° will furnish a homogeneous system of 180,000 proper motions down to about the 11th photographic magnitude and will make possible the full utilization of numerous older catalogues, especially of the Carte du Ciel. The annual proper motion probable errors derived only from a comparison with the AGK2 (1930) will be $\pm 0''.005$. In its revised form (see 1955 *I.A.U. Transactions*, p. 119) the AGK3 R involves 10 transit circle observations of each of about 21,000 reference stars for the Bergedorf plates which are $5^\circ \times 5^\circ$ and have a plate scale of 1 mm = $100''$. Twelve observatories will co-operate in this ambitious plan and one might expect probable errors of the order of $\pm 0''.10$ in the final positions. The number of stars per square degree is 5 times larger than for our 70,000 stars. The appropriate brighter star positions could usefully be incorporated into a revised GC.

It is appropriate to ask the following question: Is it feasible to increase markedly the accuracy of the photographic astrometry? The answer would seem to be in the

affirmative. The question is more "How most economically?" rather than "Is it feasible?" The one sure and quick way to increase accuracy—with a minimum of measuring effort—is by increasing focal length. The Bergedorf, Yale and Cape scales are all about 1 mm = 100″, corresponding to a focal length of 81 inches. What difficulties would be encountered if focal lengths 3 to 5 times longer should be used? These difficulties would have to do with lens optics, measuring machine and plate size, more numerous plates and reference stars. Consider an astrograph with a focal length 3 times as long (equal to 243 inches) using 24 × 24-inch plates. A 5° × 5° field is possible if a 21 × 21-inch area is used. If we assume that positions can be measured to ± 1·5 microns—this requires the best optics and superior measuring equipment—a probable error of ± 0″05 would be expected for a single measured image, provided seeing errors were negligible. If one used a 50 per cent overlap in both right ascension and declination and reversed the telescope at each observation (two images per plate) one might expect ± 0″02 for the mean of 8 images; this high accuracy would be inherent in the observational material but would be attained at a later date only when better reference star positions became available. A repetition 60 years later would provide annual proper motions with probable errors of ± 0″0005 and thus would provide "precise" motions for perhaps 99 per cent of our 70,000 stars. The measuring programme would involve measuring machines slightly larger than that at IBM (LENTZ and BENNETT, 1954) and the new Lick machine (personal communication from Dr. VASILEVSKIS); both of these will take 17 × 17-inch plates. The plates would not need to be measured both direct and reversed because of the observational technique employed and if simultaneous measures in both co-ordinates were made as envisioned in the Lick machine, the measuring programme for a hemisphere could be completed in 3 years. If the smaller 17 × 17-inch (3°·5 × 3°·5) plates were used, twice as many plates would be needed to cover the whole sky; twice as many reference stars and plate solutions would also be necessary. The increased number of reference stars required is perhaps the most serious objection to this smaller plate size, inasmuch as plate reductions would be an electronic computer job. There would be advantages in questions of plate handling, plate distortions, and in requirements on the optics.

The problem of obtaining superior optics has, historically, been a most serious one. The later Yale Zone observations were definitely improved (SCHLESINGER and MISS BARNEY, 1939) with the introduction of better objectives, a recent Yale 8-inch Ross lens needed aperture reduction to 4 inches (MISS BARNEY and VAN WOERKOM, 1954) and the Cape lens (STOY, 1954) was not only stopped down but the plate area was reduced from 5° × 5° to 4° × 4° as well. A new design effort would undoubtedly increase optical performance but there are two additional approaches possible. A photovisual rather than a photographic objective would greatly relieve chromatic aberration troubles, would reduce chromatic magnification effects, would increase the quality of the grating images of the brighter stars and would greatly reduce differential refraction effects for stars of varying colours as well. A reduction in objective size would also help. A 3-inch photovisual telescope would require only 1½ to 2 minutes of observing time for measurable images on 8·5 visual magnitude stars, and about 12 minutes for stars 2 magnitudes fainter; future emulsions will be faster and perhaps even smaller objectives can be used.

The advantage of having such a positional network of 70,000 stars is that, with proper supplementation at high galactic latitudes, it may be used as a precision

reference network for fainter stars. The increase in numbers of stars as we go fainter would point to an improvement in accuracy by increasing focal length, rather than in multiple observations. For a focal length of 365 inches and for 17×17-inch plates ($2^{\circ}5 \times 2^{\circ}5$) we might expect probable errors of the order of $\pm 0''025$ for the mean of two images from overlapping plates. The problem of good seeing might well be a critical one, as would also be the problem of adequate reference stars for this small angular area. Larger plates giving $3^{\circ}5 \times 3^{\circ}5$ (24×24-inches) might be needed. The observing times for a 3-inch lens would be about 15 minutes for stars of visual magnitude 10·5; the measuring programme would be twice as lengthy as for the 70,000 star programme—still not impossibly long. A close control on magnitude effects might be provided through the use of a grating giving images 2 magnitudes fainter; and additional shorter exposures on the same plates should be considered. The total observing time for such a hypothetical programme would be lengthy and laborious enough so that a site in an excellent climate with good seeing for each hemisphere is indicated, as is photoelectric guiding for the many moderately long exposures. The primary purpose in obtaining such accuracy would be in making possible precision proper motions for astronomers at some not too distant later epoch.

3. Magnitudes and Multicolours

As Whitford (1956) has pointed out in a paper on "Photoelectric Astronomy", the present situation is incongruous in that "the simplest and most fundamental datum about a star, its brightness, is often the least reliable entry in the catalogue". So much so that if we attempted to choose today the 70,000 brightest stars we might well misplace at least 5,000 of them. Many, if not most, present-day magnitudes are so poor that they are not quite good enough to use for setting the gain step of a photoelectric amplifier. The Johnson-Morgan (1953) U-B-V system was adopted as the standard three-colour system by Commission 25 at the 1955 Dublin meeting of the International Astronomical Union. The three-colour method has proved its worth, especially for early-type star investigations and, for example, has been used with notable success by Johnson (1957) in investigations of galactic clusters and by Hiltner (1956a) in a detailed investigation of 1,259 O- and B-stars. Adequate standards on this system are not as yet available in the southern sky. If we would extend this system to our 70,000 stars, and such extension would seem to be of first importance, we come face to face with a critical time problem which is also present in the photoelectric determination of spectral types, luminosity classes, radial velocities and polarizations. If one makes two observations of twenty seconds each in each of three colours on a star, and one such on the adjoining sky in each colour, and needs two minutes to shift from star to star, this works out to be a rate of 12 observations per hour. A standard star should be observed perhaps once every two "list" stars, so that the 12 observations are effectively reduced to 8 per hour; if more colours than 3 are observed, the number of observations per hour is correspondingly reduced. We might expect an average good location—such as at the McDonald or Lick Observatories—to yield, at best, 1,300 photoelectric hours each year with a consequent 10,400 "list" star observations a year, provided that one telescope were devoted entirely to this problem. At this rate and for a total of 2 observations per star it would take a *minimum* of 7 years to complete half the sky. This is perhaps not too impossibly long to devote to observations of such fundamental importance,

but any speed-up in observing methods would be most desirable and would make possible a useful extension to fainter magnitude limits. The problem has been discussed by the writer in more detail elsewhere (IRWIN, 1955) for the case of the 136,000 brightest stars. If the best cloud-free sites in *both* hemispheres were chosen, and if the colours were observed in simultaneous fashion by means of beam-splitters or by defining slits in the focal plane of a spectrograph, and if modern telescope setting techniques were to be used to reduce time between stars to less than a minute, then even the larger programme, with diligent application, could be completed in something less than ten years. The observational accuracy would be of the highest, not only because the sites would be chosen with this in mind, but also because the colours would be observed simultaneously; further, the tie-in to a fundamental photometric system would be unusually rigorous. Simultaneous observations from both hemispheres would provide a useful check on the vagaries of atmospheric extinction. The aperture of the telescope need be 30 to 36 inches at most, although a smaller reflector of 20 to 24 inches would be useful for setting up a network of standards. Such standards should be numerous, well spaced over the whole sky and with a common zero point, not too bright or too faint, should include the widest possible colour range and should be well tested for invariability. A careful selection of perhaps 4,000 stars mostly in the sixth to seventh magnitude range would seem to be adequate, with perhaps as many as 20 observations per star necessary to provide both high accuracy and as a test of invariability.

If such a large programme were to be attempted, serious consideration should be given to the addition of one or more colours and almost certainly a colour in the infrared at about 8,500 Å would be helpful. A star's colour depends not only on its temperature, but also on its luminosity, the amount and nature of interstellar reddening, the star's chemical composition and perhaps even its population type. A fourth colour would be of prime importance in helping to disentangle these complex effects. The U-B-V system does not, of itself, separate giants and dwarfs in the numerous early K's, for example, and KRON and his co-workers (KRON and SMITH, 1951; KRON, GASCOIGNE and WHITE, 1953 and 1957; KRON, 1954) have shown the much greater usefulness of and R – I colour index in investigations of late-type stars. JOHNSON and SANDAGE (1956) have shown that ultraviolet anomalies in the stars in Messier 13 make identification of sequences dubious at this time; these anomalies probably carry over into the blue colour region to a limited extent and indicate the need for at least one other colour longward of the visual region. Finally, an infrared colour would help to establish magnitude scales necessary to any proper discussion of objective prism investigations in the infrared with Schmidt telescopes. The main disadvantage at the present time is that the best infrared cells are down in quantum efficiency by a factor of from 20 to 40 as compared to the blue cells although this disadvantage is partially balanced by the greater usefulness of an infrared colour for the redder stars and their larger energy output in this spectral region. There are indications (SOMMER, 1955) that more efficient infrared photocells, or perhaps even multiplier phototubes, are on the way. Whether or not a red colour or a second ultraviolet colour is needed and could be efficiently included in a long observing programme should also merit consideration before such a programme is started.

Our proposed multicolour programme would provide exceptionally accurate magnditues and colours that would be of use in a wide variety of future problems, many of which cannot yet be visualized. The combination of this data with other

kinds of data such as accurate spectral classes, luminosities and velocities would be especially enlightening. Magnitude and colour data alone or combined with other kinds of data would yield a detailed knowledge of the extent, amount, and nature of interstellar reddening and absorption; it would give us large numbers of excellent photometric distances, even for distant stars; it would provide a superior network of magnitudes and colours for the whole sky; it would lead to the discovery of many new variables, unusual and composite stars and in general, would open up many new lines of research. Given two observatories, at proper locations in each hemisphere, each equipped with a 24-inch and a 36-inch reflector plus appropriate accessory equipment, then it is clear that a most ambitious programme could be carried out in a relatively few years. Modern data-processing machines would be useful in selecting programme stars, in calculating modern positions needed for telescope pointings, in performing lengthy reductions and in setting up the catalogue for final publication. Their use would make possible an early publication of the results and would remove much of the computational drudgery and, in general, the use of such computers makes feasible lengthy programmes of the type being considered that would otherwise be almost out of the question.

4. Spectral Types and Luminosity Classes

The presently accepted two-dimensional spectral-luminosity classification system is the MK system (Johnson and Morgan, 1953). No attempt will be made here to review the numerous published investigations, mostly by University of Chicago astronomers, which give data based on this system. Its refinement and extension to our 70,000 stars would be especially fruitful if magnitudes and multicolour data were also available for these same stars. There have been a number of recent successful attempts to increase the precision of both spectral and luminosity class determinations, and four of these are described in Volume 2 of *Vistas in Astronomy* (Chalonge, 1956; Strömgren, 1956a; Petrie, 1956; Heard, 1956). Of these four techniques only Strömgren's involves photoelectric observations at the telescope, and because of our serious observing time restrictions, is therefore the only one of these techniques that offers possibilities of success in this mass-production problem. The method involves photoelectric photometry with narrow-band interference filters. An index l is obtained from filters centred on and on each side of Hβ; an index c is obtained from filters centred at 3,550 Å, 4,030 Å and 4,500 Å. The indices are insensitive to interstellar reddening and somewhat insensitive to atmospheric extinction. Linear expressions for both B-V and M_v are established from nearby unreddened stars with good parallaxes. For G and K stars the Hβ absorption line and the Balmer discontinuity are no longer strong features of the spectrum so that the K-line, the G-band and cyanogen absorption near 4,200 Å are used. The spectral coverage is not as yet complete, gaps existing at B9–A2, G0–G6 and stars later than K5. A rather complete discussion of the method may be found in Strömgren (1956b) and the following data are taken from Table III of this paper. For one observation only, and including effects of cosmical scatter, the estimated probable errors in M_v are: $\pm 0^m3$, B0-B8; $\pm 0^m12$, A3–A8; $\pm 0^m17$, A9–F1; $\pm 0^m2$, F2–F9 and $\pm 0^m5$, G7–K5. The probable error of a spectral classification averages about a fiftieth of a spectral class under the same conditions.

If the error in M_v is the dominant source of error in the determinations of distances

from absolute and apparent magnitudes and colour excesses—as would generally be the case—the expected probable error in the distances from the above numbers would range from ±6 per cent for the A3–A8 group up to ±21 per cent for the G7–K5 group. These errors can be reduced by (1) taking two or more observations; (2) using more extensive and definitive calibrations—and these would be available in a large observational programme; (3) using a 15 Å half-width interference filter, rather than a 35 Å filter—and this is now being done; (4) using other strong spectral features as additional classification indices. For example, the K-line should be most useful in the G0–G6 gap and perhaps in the F's and A's. The sodium D-lines should be useful as well as Hα, although current phototube sensitivities at this latter wavelength are usually down by more than a factor of 10. For stars later than K5, photoelectric spectral classification should be possible with high precision using appropriate line and molecular band features such as Ca I at 4,226 Å and the TiO bands. Insofar as sensitivity is concerned it has been possible to go to 11th magnitude easily with the McDonald 36-inch. One would expect up to a 4-magnitude loss for a 15 Å band width as compared with the wide band widths normally used in photoelectric photometry. An $8^{m}5$ star would effectively become a $12^{m}5$ star; this can be observed with the requisite precision with a 36-inch, especially if observations were to be integrated for times up to a minute.

Another approach, which would perhaps be a more sensitive and precise modification of the interference filter method, would be to use a series of appropriate slits—each feeding a phototube—in the focal plane of a spectrograph. This should be a slitless spectrograph with large angular dispersion and a large collimator beam in order to minimize guiding errors; a coudé type spectrograph would seem to be most suitable in meeting these specifications and also because of stability and space considerations. Such a spectrometer (for Cassegrain operation however) is being built and tested at the Yerkes Observatory. Eight narrow colour regions are observed simultaneously and eight phototubes and eight pulse-counting units are employed (see, however, WEITBRECHT, 1957b). Our conclusions are that: (1) while additional future investigations are undoubtedly necessary, enough has been learned to indicate that very superior spectral and luminosity classes can be obtained over the entire spectral range (with the probable exception of a small percentage of peculiar or unusual stars); (2) a 36-inch reflector with a coudé spectrograph and with a number of narrow colour bands observed simultaneously would be adequate at $8^{m}5$ for observation times of the order of one minute; (3) parallaxes could be derived for most stars with percentage accuracies of ±5 to ±10 per cent; (4) the computing and reduction problem is a relatively simple and uncomplicated one, with no additional lengthy laboratory measurements necessary.

The STRÖMGREN technique is a real "breakthrough" and with further refinements and extensions, needs to be vigorously exploited. Such exploitation on a large scale is possible only with the construction of new and somewhat elaborate instrumentation placed in optimum locations in each hemisphere. Given such instrumentation and data-processing aids, a programme involving two observations on each of 70,000 stars could be completed in four or five years. The instrumentation and observing techniques could be made strikingly similar to that used in the magnitude and multicolour programme. The technique can be used in supplemental fashion to provide a wealth of valuable calibrations. For example, a few A3–A8 stars observed in each of all galactic clusters out to 1,000 parsecs would provide—along with the

requisite U-B-V photometry—excellent distance moduli and hence luminosities for large numbers of highly luminous stars. This would involve narrow-band photometry to 12th or 13th magnitude and this is possible if long observing times are used.

5. Parallaxes

The latest Yale Parallax Catalogue (Miss Jenkins, 1952) contains trigonometric parallaxes for 5,822 stars of which only about 1,760 are brighter than visual magnitude 8·5 and nearer than 50 parsecs. The vast majority of our 70,000 stars are over 100 parsecs distant and by no stretch of the imagination would we wish to embark upon a parallax observing programme for even 5 per cent of this group. A good trigonometric parallax requires at least 20 plates and Strand (1956) has rightly urged that selected stars be reobserved with a series of 50 plates or even more per star. Although the long-focus telescopes used in photographic astrometry have objectives of moderate size, the inclusion of a section on parallaxes in this paper is justified only because excellent parallaxes can be photometrically *derived*, given proper calibration, through the well-known equation: $M_v = m_v + 5 + 5 \log p - A_y$, where A_v is usually taken to be about 3·0 times the colour excess on the B–V system. One of the main justifications for discussing possible improved observational data for 70,000 stars relates to the tangential velocities—and space velocities—of this group; a knowledge of the parallax is necessary to convert the observed proper motions to tangential velocities. Our of this hoped-for mass of accurate basic data should come both calibrations and knowledge, from which one might derive photometric parallaxes to better than 10 per cent—perhaps 5 per cent in favourable cases—for all but the peculiar, or unusual, or most luminous, or most heavily reddened stars.

Fundamental calibration data for photometric parallaxes will also undoubtedly come from improved trigonometric parallaxes, although it is not true, as has been asserted (see Irwin, 1957b), that all estimates of celestial distances are based on trigonometric parallaxes. For example, if one examines the parallaxes of the seven nearest cepheids as listed by Weaver (1956), namely: Delta Cephei, Eta Aquilae, DT Cygni, Zeta Geminorum, FF Aquilae, X Sagittarii and RT Aurigae, the absolute trigonometric parallaxes as given in the Yale Catalogue are, respectively: $0''.005 \pm 5$, $0''.005 \pm 5$, none given, $0''.004 \pm 6$, $-0''.007 \pm 5$, $0''.028 \pm 11$ and $0''.006 \pm 4$, with a weighted mean of $0''.003 \pm 2$ (half a chance of being correct over a range of a factor of 5 provided the systematic errors are negligible!). For comparison, Weaver's parallaxes for these same stars, based on the period-luminosity law and the new distance scale, are, respectively: $0''.0032$, $0''.0028$, $0''.0026$, $0''.0023$, $0''.0023$, $0''.0022$ and $0''.0021$, with a mean parallax of $0''.0025$. If one makes the very crude assumption that each cepheid in question has a tangential velocity of 20 km/sec—a mean value consistent with the radial velocity data—and uses the "raw" total proper motions (uncorrected for solar motion) in the equation $p = 4{\cdot}74\,\mu/T$, the calculated parallaxes are, respectively, for these seven cepheids: $0''.0029$, $0''.0032$, $0''.0010$, $0''.0016$, $0''.0020$, $0''.0021$ and $0''.0031$, results in surprisingly good individual agreement with Weaver. Two conclusions follow: (1) despite laborious efforts to derive the above trigonometric parallaxes such parallaxes are useless in this fundamentally important calibration both individually and in the mean; (2) total proper

motions give surprisingly good individual parallaxes, far better than individual trigonometric parallaxes for such distant stars.

The problem could well be turned around and stars at known large photometric distances would be valuable in checking small systematic errors in the trigonometric parallax system, errors that are by no means well known (see, for example, SCHILT, 1954). More to the point, accurate photometric distances would enable parallax observers to sort out efficiently those stars in need of observing or reobserving. It may well be that all or almost all stars with accurate photometric parallaxes less than 0".02 should simply be ignored by parallax observers as not worth the effort. Over 3,000 stars, or more than half of those in the Yale Catalogue, have observed trigonometric parallaxes less than 0".02. More and better parallaxes are obviously needed, especially for many nearby faint stars such as the white dwarfs and the subdwarfs; most of these stars are so faint that they can only be reached by large reflectors blessed with good skies.

Our long-focus astrometric refractors are old instruments, having seen many decades of useful service. It would be surprising if the performance of most of them would not be markedly improved by the installation of new driving gears of the best modern design. A periodic error of even 0".1 amplitude is large if one is interested in systematic errors of the order of 0".001. A different and possibly fruitful approach would be to design a mechanical-electronic gadget that would work off the driving shaft or gears and which would effectively compensate the systematic driving errors by corrections to the driving motor rate. The driving errors of a telescope can be nicely exhibited by means of a photoelectric photometer "looking" at an out-of-focus stellar image which is being bisected by the edge of the diaphragm hole. Another and perhaps a necessary additional approach is the use of a photoelectric guider. Such a device has been built by WEITBRECHT (1957a), for example, and works effectively, using 4 photomultipliers, on stars as faint as 9th magnitude; the full $18\frac{1}{2}$-inch beam of the Dearborn Observatory refractor is used.

We might expect our best calibrations, however, to come both from more accurate moving cluster data and from really thorough investigations of a hundred or more galactic clusters with large reflectors. The basic calibration for the distances of the latter would come through knowledge of the H–R diagram correlated with both moving cluster and trigonometric parallax data, with some additional help from nearby individual stars of moderately high luminosities and measurable trigonometric parallaxes.

6. RADIAL VELOCITIES

The General Catalogue of Stellar Radial Velocities (WILSON, 1953) lists the velocities of 15,106 stars and about 12,500 of these are brighter than visual magnitude 8·5. This Catalogue is the culmination of a half-century of strenuous effort in this field, especially by observers at the large Pacific Coast observatories. Despite such efforts, our knowledge of the radial velocities of stars fainter than 7th magnitude is extremely fragmentary. Stellar spectrographs are so slow and inefficient that the mass determination of accurate radial velocities of 70,000 stars is simply out of the question if conventional methods are employed. Three years ago the writer (IRWIN, 1955) pointed out that new observational techniques might change the whole picture and make it possible to obtain 50 Å/mm spectrograms of 9th or 10th magnitude

stars in a minute or two, using a moderate-size reflector. It was suggested that LALLEMAND's technique might be developed to a practical point and applied to the problem but as H. W. BABCOCK (1955) pointed out, the application of image tubes to precision spectroscopy appears remote. He and FELLGETT (1953, 1954a, 1954b, 1955) independently emphasized that the possibility exists of determining radial velocities by a radically new technique in which a single photocell is used to measure, as a function of displacement in the wavelength-direction, the proportion of the dispersed star-light which is transmitted by a suitable template placed in the focal-plane of the spectrograph. This method would utilize the high quantum efficiency of modern photo-cathodes. Dr. FELLGETT's idea arose from an analysis of radial-velocity measurements from the point of view of information theory, and he has suggested the name "radial-velocity photometer" for an instrument working in this way. Dr. BABCOCK's proposal is that his "stellar template magnetometer" might be modified to give stellar radial-velocities. This instrument is really the outgrowth of the highly successful solar magnetograph (BABCOCK, 1953) which operates on just two spectrum lines and has been modified in such a way as to give solar disk velocities with accuracies of the order of $\pm 0{\cdot}01$ km/sec. The template is normally a contrasty (negative) spectrogram of a star of the appropriate spectral type and is slowly and accurately traversed along the spectrum. The light is transmitted by the template into a phototube with the aid of a condensing lens. An a.c. method, involving a small-amplitude oscillation of the optical spectrum, is used to locate the position of coincidence with the "live" or optical spectrum of the star and this yields the radial velocity. The equipment has been installed at the 200-inch coudé spectrograph and illustrative results have been obtained on 5th magnitude stars (personal communication).

Such a device permits utilization of the high efficiency of the photoelectric surface and has the added advantage that the effect of numerous spectrum lines are summed in one indicator reading; this is exactly what is wanted. Before any large programme is initiated, however, many questions need to be answered; some of these have been examined in FELLGETT's above-mentioned papers.

Exactly what kind of template is most efficient at each spectral class? How many templates are needed to cover the entire spectral range? What procedure is necessary in the case of double-line stars, peculiar stars, composite stars and emission objects? How long a section of the spectrum can be used for a given phototube? What is the effect of a changing camera scale or a scale change due to a large velocity? Would it be advantageous to use a number of phototubes along the length of the spectrum? How is the comparison spectrum to be managed most effectively?

It is not at all clear what accuracies might be realized at different magnitudes except to say that they should be at least equal to those obtained by photographic techniques at comparable dispersions and might be markedly better. A dispersion of 40 Å/mm should give radial velocities of most stars to $\pm 1{\cdot}5$ km/sec. However, the efficiency of the template spectrograph depends on the angular dispersion and the diameter of the collimated beam, rather than on the linear dispersion employed, which might be fixed once and for all at some convenient value. Here again, as in the case for spectral and luminosity classifications, one would suggest a coudé spectrograph with its high efficiency, stability and ease of operation.

A sample calculation is enlightening. If we should have a 36-inch reflector feeding a coudé spectrograph similar to that at the 200-inch (BOWEN, 1952), except for a

collimator beam 6 inches in diameter rather than 12 inches, we could obtain a dispersion of 39 Å/mm with an 8·4 inch focal-length camera at an f/d of 1·4; the 10,000 lines/inch grating is blazed in the third-order violet. The stellar image diameter, with the slit removed, at the spectrograph plate (or template) in a direction parallel to the dispersion would be 0·013 mm for an average seeing image diameter of 2·5 seconds of arc. Inasmuch as Δ, the width of the slit image permissible at the plate, is usually set equal to 0·02 mm, this would suggest that we need not use a slit at all, even with somewhat poorer seeing. This would have two advantages; (1) no loss of light at the slit and (2) small fluctuations in the image intensity. However, misguiding by 1 second of arc would correspond to an error of 14 km/sec at 4,500 Å and it is problematical whether guiding to better than 0·1 seconds of arc is possible as a routine thing. An automatic guider would be useful, perhaps necessary, slit or no slit. Although the total image intensity would be relatively constant in the case of a slitless device, the short-time fluctuations in the observed radial velocity due to bad seeing might be discouragingly large. Although a slitless device might be desirable for faint stars, a narrow slit admitting 40 per cent of the light—or less—would seem to be the answer. The fluctuating intensity of the light transmitted by the slit is not a critical problem in this a.c. method. The camera focal length would more conveniently be 4 or more times longer than as suggested above inasmuch as there are no particular advantages with the photoelectric approach in compressing the spectrum so much; and there are some obvious disadvantages with such a fast camera. It would be advantageous in the observational routine to use only one template on any given night or half-night and observe only stars of one spectral class for long stretches of time. Some advantage might also be had in systematically observing bright standards every three or four observations as is usual in normal photoelectric routine. A small percentage of our 70,000 stars might well give anomalous velocities because of spectral peculiarities.

If we estimate a magnitude loss at the slit, a $1\frac{1}{2}$ magnitude loss at the template and perhaps half a magnitude loss at the grating, then our 8·5 magnitude star would be comparable to an 11·5 magnitude star in normal photoelectric colour work. High accuracy at this magnitude is normally not a problem with efficient phototubes, even for observations of a fraction of a minute's duration. We are dealing here, however, with a small percentage effect which is being observed differentially; the primary source of error should be due to the statistics of the photoelectrons. A larger telescopic objective would therefore give slightly greater accuracies at the faint end of things, and because a narrower slit could be used, some improvement also for the bright stars. More definitive performance answers can be given when such a device has been perfected and thoroughly tested.

Tentatively, however, we can conclude that an 8·5 magnitude star can be observed by these techniques to our required "precision" standard of $\pm 1{\cdot}5$ km/sec, for a one-minute observation; it may turn out to be somewhat better. The developing and perfecting of such a revolutionary type of instrument offers a special challenge in that it would make radial velocity determinations of the requisite accuracy for our 70,000 stars an observational effort comparable to that necessary for either magnitudes and multicolours, or spectral and luminosity classes, or polarizations for these same stars. The primary purpose in such a programme would be in the determination of the magnitudes and directions of the space velocities of these stars; of secondary interest would be the discovery of many new high-velocity stars and

stars of variable velocity, and also the possibility of determining many high quality statistical parallaxes.

7. Polarizations

The discovery of interstellar polarization is of such recent origin that a section on it could not have been included in a paper similar to this written ten years ago. Hiltner (1956b) has pointed out: "At present, polarization measurements provide the only access to the observational characteristics of the magnetic fields which must play an important rôle in the structure of the Galaxy". He further states: "The need for accuracy in most measurements of stellar polarization cannot be over-emphasized. Often the quantity to be measured is only a few hundredths of a magnitude and not infrequently smaller than the precision of the instrument employed". Precision polarization data for large numbers of stars are needed for a variety of reasons and would be especially useful when combined with data on magnitudes, colours, distances, and spectral and luminosity types. Inasmuch as Hiltner's paper is a useful summarizing discussion of the problems and possibilities in this exciting new field, no further discussion of this sort need be attempted here. Polarizations for 551 early-type stars have been published by Hall and Mikesell (1950), for 841 stars by Hiltner (1951a), for 450 high-luminosity stars by Hiltner (1954), and polarizations, U-B-V colours, and spectral and luminosity data for 1,259 O- and B-stars by Hiltner (1956a). There have been a variety of other, less extensive observational investigations but it is probable that at least 95 per cent of our 70,000 stars have not been observed for polarization.

The most accurate observational technique is Hiltner's (1951b) split-beam differential method in which the seeing noise is "compensated". A Wollaston prism separates the perpendicular planes of vibration and the outputs of two moderately well-matched phototubes are compared in a differential amplifier. In practice, μ Herculis, a star of visual magnitude 3·4, was observed with an accuracy of $\pm 0^{m}00016$ (p.e.), corresponding to a probable error of $\pm 0{\cdot}007$ per cent in the polarization; this was for a 4-minute observation with the 82-inch McDonald reflector. Further observational results, using this technique, have not been published to the writer's knowledge. The method works most efficiently for brighter stars and is of equivalent accuracy with Hiltner's d.c. method for stars at about tenth magnitude with the 82-inch. If we assume that phototubes twice as efficient as those used by Hiltner can be obtained, the performance data for 1-minute observations with a 36-inch reflector can be calculated on the basis that the errors are of statistical origin only. The calculated probable errors expressed in per cent are: $\pm 0{\cdot}025$ at 3·5 visual magnitude, $\pm 0{\cdot}079$ at 6·0 visual magnitude, and $\pm 0{\cdot}25$ at 8·5 visual magnitude. These accuracies are at least comparable with anything that has been published and should be sufficiently good for most purposes. This is one type of observation, however, that could be improved by the use of a larger telescope or a longer observing time, or both, but at this time we do not know how important such super-accurate polarizations might be. Quantum efficiency of the phototubes used would be of prime importance in attaining highest accuracy.

8. Conclusions

A brief recapitulation of the preceding discussions would be as follows. A revised and extended Boss General Catalogue of 70,000 stars would yield proper motions

perhaps four to five times as precise as now available, this improvement coming about both from a discussion of published observations and from observations that will shortly be, or can shortly be, made. The proper motion accuracies needed for stars of high luminosity are so high, however, that every effort must be made to improve our astrometrical techniques for these brighter stars by substantial amounts as soon as possible. There are reasons to think that such "breakthroughs" are possible, perhaps by using phototubes on transit circles in the case of fundamental observations and by using wide-angle photovisual astrographs of much longer focal length in the case of differential observations. Of perhaps even more immediate concern would be the obtaining of multicolour data for these 70,000 stars; these basic data are not known for even the majority of stars brighter than sixth magnitude. One additional colour in the infrared—at least—should be added to the JOHNSON-MORGAN U-B-V System. The STRÖMGREN techniques, suitably extended to stars of all spectral classes, would give spectral types and luminosities of a new order of precision. These, combined with the multicolour information and with appropriate calibration, would yield very high quality photometric distances. The FELLGETT-BABCOCK photoelectric template technique should be applicable to the mass determination of high-quality radial velocities with short observing times per star and should greatly simplify and shorten reduction procedures. The HILTNER split-beam technique would give accurate polarizations at 8^m5 and would yield accuracies ten times as large for stars 5 magnitudes brighter. Only in this kind of observation and in the radial velocity case might we expect any considerable improvement in accuracy by going to reflectors larger than 36 inches in diameter.

The obvious conclusion arising out of these discussions is that, despite lengthy and strenuous past efforts, our basic knowledge of the brighter stars is extremely limited and could be—and should be—vastly improved; and that such improvement in knowledge could be obtained, with proper effort, in the next decade or two. The reason for this is, for the most part, due to the development of efficient multiplier phototubes which, when substituted for the photographic plate, can simultaneously speed up and sharpen up most of the kinds of observations that have been discussed. The substitution of the phototube for the photographic plate brings with it a fundamental shift of emphasis in the sense that most of the work must be done at the telescope at night—star by star—rather than with some measuring machine during the day. This in turn requires the location of observatories in regions as free of clouds and haze as possible, that is, in dust-free arid or semi-arid regions. Such a "photoelectric observatory" has been discussed by the writer (IRWIN, 1952), with special emphasis on possibilities in the southwestern United States; but other places in the northern hemisphere such as in Spain or in North Africa or in the U.S.S.R. might well qualify. The southern hemisphere is even more important (IRWIN, 1957a; IRWIN and VELGHE, 1957) and there are excellent site possibilities in north-central Chile, South and Southwest Africa and perhaps in Australia and New Zealand. Quality of seeing is only of minor importance in photoelectric observations of bright stars, or even of stars some 5 magnitudes fainter than our limit of 8^m5, and this fact drastically simplifies the problem of site selection.

A really good site might well have 200 to 250 cloud-free, dust-free nights a year and we might therefore expect some 2,000 "photoelectric" hours a year as a conservative minimum. In each of the four kinds of photoelectric observations that we have considered, proper instrumentation should keep the observational time per

star to less than $1\frac{1}{2}$ minutes; and with one minute to shift from star to star, some 24 observations an hour are possible. One third of these might be "standards" so that a total of 32,000 "list" star observations a year would be possible with proper application and an adequate number of observers and assistants. At two observations a star, one hemisphere would be "cleaned up" in just over two years; a more conservative and realistic estimate might be 4 or 5 years of observing in each of the four categories in each hemisphere. This is a surprisingly short time and is strongly indicative of the feasibility of completing such a large observing programme. A typical photoelectric observatory might have three instruments: (1) a 24-inch reflector for magnitude and colour standards; (2) a 36-inch reflector for magnitudes and colours along with polarizations; (3) another 36-inch reflector for spectral and luminosity classes along with radial velocities. Such an observatory could be established and put into full operation in a much shorter time than is necessary to complete a large reflector installation. No lengthy seeing tests are necessary and 36-inch reflectors are relatively easy to construct and figure.

So much pioneering work would be possible at such an observatory—for example in galactic cluster and variable star investigations—over and above the demands of basic observational programmes even larger than those outlined here, that a photoelectric observatory could easily "stand on its own feet" and might well be of decisive importance in guiding observational programmes at larger observatories and in making observations with our few, very expensive reflectors both more efficient and more meaningful. There are no photoelectric observatories in the world today of the type envisaged; nor does the writer know of any plans for such an observatory. Yet the research possibilities for one are so great that it should be only a matter of time before some are established. Photoelectric sites might be useful for astrometric research also, not only because only the nights of top seeing chosen from a large number of nights need be used, but also because either fainter stars could be observed or shorter exposures could be made—at a given magnitude level—as compared with less favourable locations with hazier atmospheres.

The photoelectric surface, with its ability to change starlight efficiently to more easily manageable and measurable electricity, should not be the only reason for a change in our attitude towards lengthy and laborious observational programmes. Modern data-processing machines employing electronic computers are now fairly common and would not only remove much of the drudgery associated with such programmes in the past but would also greatly speed up publication of results. Furthermore, such computers would be of inestimable value in setting up an observing programme involving tens of thousands of stars, both in selecting the stars and in providing modern positions accurate enough for circle settings to be made.

The point of view taken here has been that all these basic data are necessary in order to improve our knowledge of stellar motions, especially in connection with problems of stellar evolution; but this is only a part of it and perhaps not the most important part. The greatest importance of such a treasure chest of basic data would almost certainly be in new discoveries or in the solution of new problems that we do not know of today at all. The technological breakthrough of the multiplier phototube would make such a large observational effort timely in the sense that, barring the possibility of a space station, further great improvements in accuracy would seem somewhat remote. It is also timely because the next few years will see the initiation into full-scale operation of a number of large reflectors such as the Lick

120-inch, the Canberra 74-inch, the European 120-inch in South Africa, the National Astronomical Observatory's 80-inch in Arizona, and perhaps something even larger that 200 inches both in Arizona and in the U.S.S.R. These large, expensive reflectors build upon and make use of knowledge obtained with smaller instruments. The cost of even a pair of photoelectric observatories would be substantially less than for the Lick 120-inch or the National Astronomical Observatory and such photoelectric observatories could be built and the four extensive photoelectric programmes completed in the time it would take to put into operation a large reflector such as the 120-inch.

Piecemeal, all these data for 70,000 stars might take a century or more to obtain, and the results would neither be as homogeneous nor as accurate as if mass production methods were to be employed for the next decade or two under optimum conditions. Astronomers are obviously going to obtain this data at some indefinite future date; it would seem to be sensible to meet the challenge at once and get on with the job, now that the tools are at hand.

REFERENCES

BABCOCK, H. W.	1953	*Ap. J.*, **118**, 387.
	1955	*Carnegie Inst. of Washington Year Book*, No. 54, p.27.
BARNEY, I. and VAN WOERKOM, A. J. J.	1954	*Yale Obs. Trans.* **26**, 5.
BLAAUW, A. and MORGAN, H. R.	1954	*Bull. Astron. Inst. Netherlands*, **12**, 95.
BOWEN, I. S.	1952	*Ap. J.*, **116**, 1.
CHALONGE, D.	1956	*Vistas in Astronomy* (Ed. A. Beer), Vol. 2, p. 1328. (Pergamon Press, London.)
DELHAYE, J. and BLAAUW, A.	1953	*Bull. Astron. Inst. Netherlands*, **12**, 72.
FELLGETT, P. B.	1953	*J. Opt. Soc. Amer.*, **43**, 271.
	1954a	*Vistas in Astronomy* (Ed. A. Beer), Vol. 1, p. 484. (Pergamon Press, London.)
	1954b	*Proceedings of the Florence Meeting.* "Problems in Contemporary Optics" of the Istituto Nazionale di Ottica, 10–15 September 1954, Firenze.
	1955	*Optica Acta*, Paris, Vol. 2, p. 9.
FORBES, G.	1916	*David Gill, Man and Astronomer*, p. 222. (John Murray, London).
HALL, J. S. and MIKESELL, A. H.	1950	*Publ. U.S. Naval Obs., Washington*, (Second Series), **17**, 16.
HEARD, J. F.	1956	*Vistas in Astronomy* (Ed. A. Beer), Vol. 2, p. 1357. (Pergamon Press, London).
HECKMANN, O.	1954	*Astron. J.*, **59**, 31.
HILTNER, W. A.	1951a	*Ap. J.*, **114**, 241.
	1951b	*Observatory*, **71**, 234.
	1954	*Ap. J.*, **120**, 178.
	1956a	*Ap. J. Supplement*, Vol. 2, No. 24.
	1956b	*Vistas in Astronomy* (Ed. A. Beer), Vol. 2, p. 1080. (Pergamon Press, London.)
IRWIN, J. B.	1952	*Science*, **115**, 223.
	1955	*Sky and Telescope*, **14**, 132.
	1957a	*Scientific Monthly*, **84**, 111.
	1957b	*Sky and Telescope*, **11**, 542.
IRWIN, J. B. and VELGHE, A. G.	1957	*Ciel et Terre*, **73**, 439.
JENKINS, L. F.	1952	*General Catalogue of Trigonometric Stellar Parallaxes.* (Yale University Observatory, New Haven).
JOHNSON, H. L.	1957	*Ap. J.*, **126**, 121.
JOHNSON, H. L. and MORGAN, W. W.	1953	*Ap. J.*, **117**, 313.
JOHNSON, H. L. and SANDAGE, A. R.	1956	*Ap. J.*, **124**, 379.
KRON, G. E.	1956	*Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, p. 39. (University of California Press.)
KRON, G. E., GASCOIGNE, S. C. B. and WHITE, H. S.	1953	*Ap. J.*, **118**, 502.
	1957	*Astron. J.*, **62**, 205.

Kron, G. E. and Smith, J. L.	1951	*Ap. J.*, **113**, 324.
Land, G.	1954	*Astron. J.*, **59**, 26.
Lentz, J. and Bennett, R.	1954	*Electronics*, **27**, No. 6, 158.
Markowitz, W.	1954	*Astron. J.*, **59**, 69.
Morgan, H. R.	1906	*Popular Astron.*, **14**, 452.
Morgan, H. R. and Lyons, U. S.	1937	*Publ. U.S. Naval Obs.*, Second Series, **14**, 81.
Morgan, H. R. and Scott, F. P.	1948	*Publ. U.S. Naval Obs., Washington*, Second Series, **15**, 115.
Morgan, W. W., Whitford, A. E. and Code, A. D.	1953	*Ap. J.*, **118**, 318.
Petrie, R. M.	1956	*Vistas in Astronomy* (Ed. A. Beer), Vol. 2, p. 1346 (Pergamon Press, London).
Russell, H. N.	1952	*Astron. J.*, **57**, 1.
Schilt, J.	1954	*Astron. J.*, **59**, 55.
Schlesinger, F.	1916	*Allegheny Obs. Publ.*, **3**, 1.
Schlesinger, F. and Barney, I.	1939	*Yale Obs. Trans.* **11**, 5.
Shapley, H.	1933	*Handbuch der Astrophysik*, V/2, 766. (Julius Springer, Berlin).
Sommer, A. H.	1955	*Rev. sci. Instrum.* **26**, 725.
Stoy, R. H.	1954	*Astron. J.*, **59**, 35.
Strand, K. Aa.	1956	*Smithsonian Contr. to Astrophysics*, **1**, 21.
Strömgren, B.	1956a	*Vistas in Astronomy* (Ed. A. Beer), Vol. 2, p. 1336 (Pergamon Press, London.)
	1956b	*Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, p. 49. (University of California Press.)
Watts, C. B. and Adams, A. N.	1949	*Publ. U.S. Naval Obs., Washington*, Second Series, **16**, 3.
Weaver, H. F.	1956	*Astron. J.*, **61**, 268.
Weitbrecht, R. H.	1957a	*Rev. sci. Instrum.*, **28**, 122.
	1957b	*Rev. sci. Instrum.*, **28**, 883.
Whitford, A. E.	1956	*Smithsonian Contrib. to Astrophysics*, **1**, 25.
Wilson, R. E.	1953	*General Catalogue of Stellar Radial Velocities*. Carnegie Inst. of Washington Publ. 601.

Multiple Colour Photometry

Gerald E. Kron
Lick Observatory, University of California

Summary

Multiple colour photometry is discussed historically. Modern photometric systems are discussed and bibliographies are given. Demonstrations of some of the properties and uses of multiple colour photometry are given, with emphasis placed on the infrared spectral region. The determination of colour temperatures by means of multiple colour photometry is discussed.

Multiple colour photometry is a natural outgrowth of simple photometry and two-colour photometry, both of which have a very long history of application to the problems of observational astronomy. The extension of photometric technique to many colours is a reasonable consequence of the often incorrect primitive philosophy of abundance: if a little is good, a lot must must be better.

Simple stellar photometry has been practiced since the time Hipparchus graded stars according to their brightness, about 130 B.C. Photometry in two colours apparently started in earnest when Secchi, in 1852, compared the colours of stars with the colours of certain emission lines in comparison spectra, a procedure that was greatly improved by Chandler who, 33 years later, employed a colour filter to reduce the subjectivity of eye estimates of colour. The first useful published multiple colour photometry, by E. Öpik who used photographic methods, followed in 1925 after the lapse of an additional 37 years. After Öpik's work, new events crowded together in the rapidly accelerating pace so characteristic of technical developments in the last few decades.

1. Introduction

Spectrophotometric investigations have shown that star light is black-body radiation of the temperature of the star, modified by certain complex absorption features. The absorption effects, of which the well known "Balmer jump" (for example, C. S. Yü, 1926) is an outstanding example, originate in the stellar atmosphere, and are characteristic of this medium. Generally, the departures from black-body radiation are small. An observer finds that a smooth relation exists between the colour differences for single unreddened stars and their temperatures, or spectral types. Consequently, colorimetric methods have long been used for detecting and measuring the effects of the selective absorption of light by interstellar material.

Generally speaking, it is the exploitation of departures from black-body radiation that forms one of the major fields for the application of multiple colour photometry. It is possible to investigate phenomena associated with stellar temperature, luminosity, space reddening, and multiplicity. Colorimetric methods have their greatest effectiveness when combined with spectrographic data. However, an interesting

field for future development lies in perfecting methods for separating stellar parameters by means of colorimetric methods alone, for colorimetry can be employed on fainter objects than spectroscopy. To gain such ends, it may be necessary to depend upon an increase in the multiplicity of the colours, for it is by such means that the observer can gain the added observational information needed to solve problems of increased complexity.

2. General Remarks on Colorimetry

The term colorimetry is reserved to describe measurements of the brightness of a luminous source in more than one discrete, reproducible, spectral region. In astronomy, the brightnesses are expressed in magnitudes from some arbitrary zero point, and a "colour difference" is the algebraic difference $(b–r)$ where b is the magnitude in a relatively short wavelength and r is the magnitude in a relatively long wavelength. The Colour Index is a specifix colour difference formed by subtracting the visual magnitude from the photographic magnitude. Spectrophotometry differs from colorimetry in being either a continuous run of intensity measures with wavelength, or at least a very large number of measures of narrow bandwidth, closely spaced in wavelength. The methods and many of the applications of spectrophotometry have been lucidly described by W. M. H. Greaves (1948). The spectral regions used in general-purpose astronomical colorimetry are usually wide—several hundred to more than a thousand angstroms—for the purpose of making the fullest use of the precious faint light, as well as for smoothing the effects of the line structure of the stellar spectra. The spectral regions can be isolated by means of the prism, the grating, or by ingenious related methods such as Tikhoff's (1916) use of the usually undesirable chromatic effects of the lens of a refracting telescope. By far the most common method, however, is by the use of ordinary glass transmission filters. Such filters give good economy of light, they are cheap, readily available, and their manufacturers control their properties so well that the reproducibility is good. Coming into use also are filters in which the optical effects of multiple-layer interference films are employed to get band-pass or band-rejection effects, with good sharp cut-off and good economy of light. All filters must be used with careful attention to light leaks in spectral regions where the light receiver may be sensitive, but where the filter transmission properties may be unknown or not carefully controlled by the manufacturer. It is sometimes necessary to combine filters of various properties to gain needed transmission characteristics.

The development of multiple colour photometry was long hindered by a lack of appropriate light sensitive devices, or light receivers as they are sometimes called. The eye is, of course, almost useless for the purpose, and the old potassium hydride photocell, though free of the Purkinje effect, had too narrow a region of spectral sensitivity to be useful for more than two-colour photometry. It was the introduction of photographic plates sensitive to wide spectral ranges, and the application of the CsO-Ag and the alloy (Sb-Cs) photosurfaces that made practicable routine multiple colour photometry. In still more recent times, the lead sulphide photoconductive cell and the germanium photo-transistor have made possible measurements in spectral regions beyond 1 micron, where transmission "windows" exist in the atmosphere of the earth, and where the photographic plate is insensitive.

The so-called "photometric systems" result from measurements in which the zero

points and scales of each colour have been carefully defined and standardized usually by means of stellar standards of colour and brightness. Such systems can be reproduced by observers with equipment different from the original, and hence observations that can be compared directly with originals can be made at later dates. A given system cannot be reproduced with absolute precision; filters, light receivers, and telescopes differ among themselves too much for this. It has been found, however, that a given observational system can be reduced empirically to a given standard system with high precision, provided certain precautions are observed. Generally, the wavelengths of the colours should be reproduced with fair precision, by employing filters and light receivers similar to those used in the original system. If photometric measures made with a CsO-Ag photocell must be reproduced, it is sometimes necessary to modify filters by experimental reference to the standards, until the observations themselves indicate that a satisfactory copy of the original system has been obtained.

3. BIBLIOGRAPHICAL

In 1925, E. ÖPIK (1925) published a list of over 100 stars whose magnitudes he had measured on a violet-blue-yellow photometric system. The observations were made with two kinds of photographic plates with TIKHOV's method. ÖPIK showed that either of the colour differences violet minus blue or blue minus yellow gave good separation of giant from dwarf stars when plotted against the Mt. Wilson spectral data. His success in making this separation by means of his colour differences alone, however, was limited.

JOHN S. HALL (1934) published a list of three colour photometry of 347 stars. Measurements were made with a CsO-Ag photocell refrigerated with solid carbon dioxide, employed first with no filter, and then with two different kinds of red filter. The three colours were not well defined, and so this pioneering effort was noteworthy mainly for calling attention to Hall's initial use of refrigeration for reducing the dark noise of photoelectric receivers, a technique of great value that is still in use.

W. BECKER (1938) gave impetus to the pursuit of multiple colour photometry when in 1938 he showed how three-colour photometry could be used on stars to separate thermal from spacing reddening, thereby paving the way for a new approach to the problem of determining colour excesses. Later, BECKER and STEINLIN (1956) showed that stars having known or equal reddening could be separated according to luminosity class, provided that they were of late spectral type, and provided that one of the three colours of the photometry lay in the red spectral region. BECKER, and BECKER and STEINLIN employed photographic photometry with transmission filters, and applied themselves to the investigation by means of colour measurements alone, of fields of stars too faint for slit spectroscopy.

HALL (1941) published a list of photometric measures of 67 stars, called by him spectrophotometry. Although the separation of the colours was obtained by means of a grating and a slit, as in spectrophotometry, the spectral regions were rather wide (about 500 Å), and the results were published as a series of measurements in thirteen discrete spectral regions from 4560 to 10,320 Å for each star, as in colorimetry. The results are of interest here because they can be considered as equivalent to multiple colour photometry of photoelectric precision with the colours delineated with sharp cuts. Besides their usefulness in themselves, HALL's results can be used

to test wide band filter systems for effects thought to be caused by the properties of the filters.

Probably the most versatile of the multiple colour systems is the six-colour photometry of STEBBINS and WHITFORD (1943). This photoelectric photometry is based upon the use of the CsO-Ag photoemissive surface employed with a series of compound glass transmission filters. The CsO-Ag photocathode must have sensitivity out to about 12,000 Å; cathodes with this property can usually be found by selection. The designers of the six-colour system adopted two principles that were new to wide band photometry. First, the filters were designed so that the effects of the Balmer discontinuity were confined to the ultraviolet filter alone, and were absent from all others. Second, the spacing and bandwidths of the filters were approximately uniform in $1/\lambda$ instead of in λ, thereby putting the spacing on a quantum energy basis, and giving the various intervals about equal physical resolving power. The effective wavelengths of the six colours range from about 3500 Å for the ultraviolet through to 10,300 Å for the infrared. The colours were isolated by means of seven filters originally, the seventh filter being for the purpose of measuring the infrared leak that all ultraviolet glass transmission filters have. During the last few years an ultraviolet filter has been used at the Lick Observatory that closes the infrared leak with an aqueous solution of copper sulphate (STEBBINS and KRON, 1957). The published observations on the six-colour system include lists totalling 409 stars (STEBBINS and WHITFORD, 1945; STEBBINS and KRON, 1956), complete light curves for five intrinsic variable stars including three Cepheids, one cluster type variable and a Beta CMa type variable. Subjects covered range from a determination of the properties of selective absorption by interstellar material to the measurement of the colour and magnitude of the Sun. A complete bibliography of published six-colour photometry is given in the table of references of the present paper. Unpublished as yet are data on the colour temperature of the Sun and stars, and new measurements of supergiant stars including many cepheid variables.

BECKER (1946) proposed a complete reform of astronomical photometry based upon a four colour system of photographic photometry. The effective wavelengths of BECKER's four spectral regions were approximately 3600, 4200, 4800, and 6400 Å, obtained by using glass transmission filters with two different kinds of photographic plates. The 4200 Å colour was rendered free of the effects of the Balmer absorption by employing a Schott GG-13 filter, as used in the six-colour photoelectric photometry and later in the three-colour photoelectric photometry of JOHNSON and and MORGAN. BECKER's published magnitudes included new measurements of an adequate list of 66 Polar Sequence stars measured in five colours, the fifth being a photovisual at approximately 5600 Å. In addition, magnitudes are given in three colours for stars in the central regions of Selected Areas Nos. 19 and 89. A list of measures on 146 stars in NGC 7654 is also given in terms of the magnitude at 4800 Å and four colour-differences. This work of BECKER's was apparently ahead of its time. It is becoming increasingly evident that enduring standards of stellar photometry will have to be based on multicolour photometry of at least four, and possibly more colours, and that in assigning wavelength regions some compromise will have to be made between the demands of astrophysics and the traditional magnitude systems.

The three-colour U, B, V photometric system of JOHNSON and MORGAN (1953), a general purpose type of photoelectric photometry, employs the very sensitive

Sb-Cs photosurface found in the 1P21 RCA multiplier and others, with glass absorption filters. The three colours are at effective wavelengths of approximately 3600, 4300 and 5500 Å. The 5500 Å magnitudes are the basis for a magnitude system adjusted to the International Photovisual. The colour differences $(U–B)$ and $(B–V)$ are made equal to zero for A0V stars and are $+1{\cdot}01$ mag. for K0III stars. The system is adequately standardized in a list of 290 stars of all luminosity classes, spectral types, and a wide range of reddening. The effects of the Balmer absorption are confined to the ultraviolet colour along by appropriate selection of filters,which makes it impossible to place the blue magnitudes on the International Photographic system without resort to a non-linear transformation. The designers of the U, B, V system therefore abandoned entirely the traditional photographic system. The U, V, B system has been used vigorously by its designers and by other observers for solving problems in many fields of astronomy, but the chief application has been to the study of the stars in clusters. Many valuable observations have been published on this system, and the complete bibliography is a long one. Chief among the published magnitudes of general interest is the original list of standards of known spectral types (JOHNSON and MORGAN, 1953) and the long list of 1,259 O and B stars by HILTNER (1956) with U, B, V photometry, polarization data, and spectral types based on slit spectroscopy.

An interesting and useful list of measures in six colours has been published by WOOLLEY, GASCOIGNE, and A. DE VAUCOULEURS (1955). These data, published in the form of magnitudes in each colour, are representative of narrow band photographic photometry, quite similar to HALL's photoelectric work. The effective wavelengths are at 4050, 4250, 4550, 4954, 5980 and 6360 Å. Magnitudes are available for 85 southern hemisphere stars. Colour differences computed from the data correlate well with available photoelectric colour differences.

A three-colour photometric system of blue, yellow, and infrared has been developed at the Lick Observatory. The system, designated by ***P, V, I*** has some features in common with the JOHNSON and MORGAN system. The yellow (V) magnitudes are adjusted to the International Photovisual, so that these magnitudes may be compared with equivalent ones by other observers without transformation. The blue (P) colour is obtained with a filter that removes effects of the Balmer absorption. The blue magnitudes are adjusted to be a compromise fitted to the International Photographic, so that the P magnitudes can be compared directly with photographic magnitudes by other observers with a minimum of systematic error. The compromise is the same as the one adopted by O. J. EGGEN (1955) in his extensive two-colour photometry. The third colour, the infrared (I), is adjusted to be on the infrared system of KRON, WHITE, and GASCOIGNE (1953), with an effective wavelength of about 8250 Å. There are as yet no published results with this system, but it has been used extensively by KRON and MAYALL in a programme of the integrated magnitudes of star clusters, now being prepared for publication. This three-colour photometry has made possible the measurement of colour excesses for globular clusters, and has given some indication concerning apparent intrinsic differences among globular clusters, especially for clusters found in different galaxies. Also, it has enabled the establishment of criteria for the separation of open clusters from globular clusters, useful when other means fail, as in distant galaxies. This three-colour system shares with other systems having a colour in the longer wavelengths the property of being sensitive to luminosity differences among the late-type stars, as

described by BECKER and STEINLIN (1956); hence, under certain favourable conditions it can be used to distinguish luminosity of stars.

4. Detached Multiple Colour Systems

When multiple colour photometric measurements are made, magnitudes in the various spectral regions are usually measured at one time, or as nearly at one time as possible, and with one light receiver. Although this method yields colour differences of highest accuracy, it is quite possible to construct multiple colour systems from observations that have been made in various spectral regions at different times with different receivers. Accuracy can be quite good if the separate measurements are made with careful reference to standards, and provided that the objects measured do not vary significantly. A good example of the use of a detached colour scheme is in the literature. In his investigation of the interstellar absorption characteristic, A. E. WHITFORD (1948) required a very long colour base line, longer than that afforded even by the six-colour system. Although the photocell employed in the six-colour photometry is not sensitive beyond 12,000 Å, WHITFORD added an observed point at 22,000 Å by making separate observations with a lead sulphide cell and combining them with regular six-colour observations. Likewise, KRON and MAYALL obtained an infrared measurement for some of the star clusters in M 31; here, magnitudes in ***P*** and ***V*** could be obtained only with a blue sensitive receiver like the 1P21. The infrared magnitudes were obtained separately with an infrared sensitive multiplier supplied by A. LALLEMAND of the Paris Observatory, thus completing the ***P***, ***V***, ***I*** three-colour system. The construction of a multiple colour system from detached measurements may not represent the best possible technique, but it is useful because sometimes it does permit the observer to make measurements, that otherwise might be impossible, and could conceivably sometimes allow the use of data already on hand as a substitute for making new observations. This possibility is illustrated in Fig. 1, where values of $(B–V)$ from JOHNSON and HARRIS (1954) are plotted as ordinates against $(V—\boldsymbol{I})$ from JOHNSON and HARRIS' values of V and KRON, WHITE, and GASCOIGNE's (1953) values of ***I***, plotted as abcissae, for some stars in common with both sets of data. Clearly, this artificially constructed three-colour photometry shows ability to separate late type unreddened giants from late type unreddened dwarf stars.

5. Some Properties of Multiple Colour Photometry

The six-colour photometry is particularly advantageous for the demonstration of the properties of multiple colour systems owing to the number of colours available, the wide range of wavelength covered, and the large and versatile lists of measurements available in the six-colour publications. In any multiple colour system, where n is the number of colours, colour differences in the number $n(n–1)/2$ can be formed. For the six-colour photometry, no less than 15 colour differences are inherent in the data. For the present demonstration we will form the two colour differences (V–B) and (R–I). The U colour will not be used, partly because of the complications of the thermal loci of stars caused by the Balmer absorption, partly because the properties of a multiple colour system containing the ultraviolet have already been demonstrated thoroughly by JOHNSON and MORGAN (1953). The colour differences here

adopted have been chosen because they give a maximum discrimination between stellar thermal and reddening loci for a system not including the ultraviolet, and also a maximum separation of the luminosity classes for the stars of later spectral types. The loci in Fig. 2 result from plotting all data on stars available in previous papers, as well as some unpublished data on supergiant stars, and a small amount of unpublished data for high velocity dwarf stars. The composite diagram is, of course, quite complex. One cannot enter into it data from just any star, and expect to interpret the results unambiguously. If one has some knowledge, however, much can be learned in addition. The interpretation for the OB stars is simple, as the thermal loci

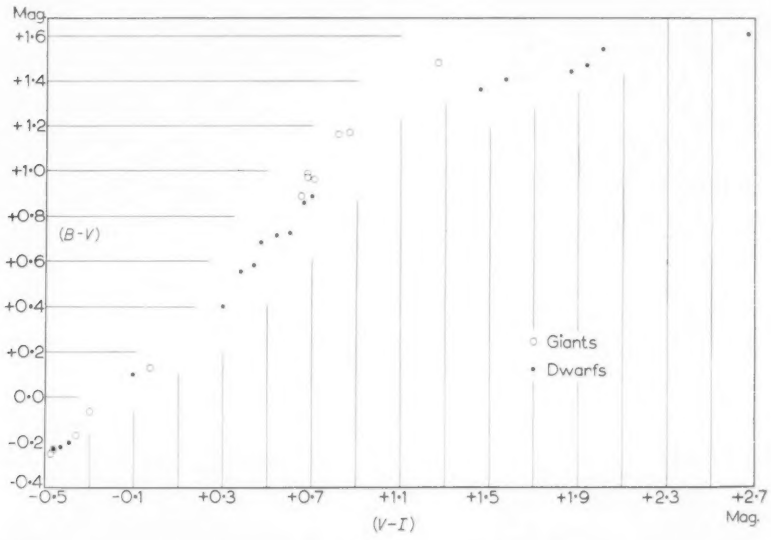


FIG. 1. Plot of three-colour photometric data for giant and dwarf stars, showing separation of the luminosity classes for stars of late spectral type.

are apparently the same for all luminosity classes. The plotted observed point may be moved parallel to the reddening locus until it lies on the OB sequence in the place appropriate to its spectral type, which must be known. The displacements on the two axes represent the colour excesses in the two-colour difference systems employed here. A calibration has shown that the (V–B) colour difference can be transformed to the base line of the International Colour Index for the transformation of colour excesses simply by multiplying by the factor 1·8. Some OB stars in the six-colour publications are common with those in HILTNER's (1956) long list that includes his derivations of colour excess, and colour excesses derived from the six-colour data by the process just described are in good agreement. Inasmuch as few known dwarf stars are appreciably reddened, Fig. 2 may be used to separate main sequence dwarf stars from the high velocity dwarfs, which have different properties, as has long been known. Velocity alone, though, is not always a reliable criterion of the more interesting physical properties that distinguish these two kinds of dwarf, and the luminosity criterion is not useful unless an accurate parallax is available. However, the good separation of the normal and high velocity dwarf branches of the curves of Fig. 2 show how well this multiple colour photometry can perform a separation,

except near the intersection of the two curves near (V–B) = 0·4 and (R–I) = 0·5. The widening of the plot for the A-type stars is caused partly by the effects of a scatter caused by varying degrees of metallicity among the dwarf stars, and may also have its origin partly in the Paschen absorption which has a small effect on the R colour of the six-colour photometry (STEBBINS and KRON, 1956).

In Fig. 3, for the sake of simplicity, only the thermal locus for supergiant stars is shown, along with plotted observed points for a variety of interesting objects. The path in this diagram of the Cepheid variable η Aql (STEBBINS, KRON, and SMITH, 1952) is plotted to demonstrate some properties. If the variable is assumed to be a normal class Ib supergiant at miniumm, then its colour excess may be determined at once by translating the observations along a reddening locus until they lie on the

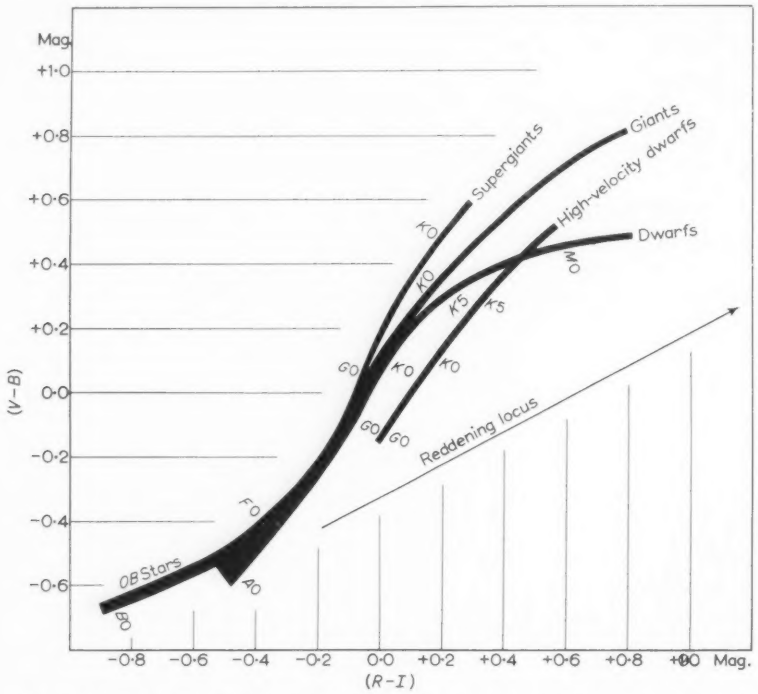


Fig. 2. Thermal loci for normal single stars on a three-colour photometric system from the six-colour photometry of STEBBINS and WHITFORD. The direction of the reddening locus is indicated.

thermal locus; multiplying the necessary displacement in (V–B) of 0·10 mag. by 1·8 gives $E = 0{\cdot}18$ mag. Plotted also in Fig. 3 are the data for the heavily reddened supergiant HD 187238, a star of spectral description K3Iab according to BIDELMAN (1957). The present analysis indicates that this star has a colour excess of 0·76 mag. The behaviour of a double star is indicated by the plotted point of the data for the well-known eclipsing binary ζ Aur. If the plotted point for this star is brought down to the thermal loci, it intersects at about F5. Unless much is known about the components of a double star, the colour plot of the type shown in Fig. 3 yields little information of value. Double stars always plot to the right of the thermal loci, so that the effects of their composite character can lead one to think they are reddened

single stars unless spectrographic data are available for revealing the true state of affairs. Very useful information can be obtained from multiple colour data however, for highly composite objects like globular clusters. Data for six globular clusters are plotted (STEBBINS, 1950) in Fig. 3. STEBBINS did not give violet measurements for the three very red clusters, but I derived approximate values for them from unpublished three-colour measures by KRON and MAYALL. Three clusters, NGC 6205, 6341, and 7089 are known to be reddened little or not at all (ARP, 1955), while the other three, NGC 6440, 6553, and 6760 are much reddened. The plot indicates surprisingly little spread from intrinsic differences among the clusters. If the displacement between the two groups of clusters is entirely caused by reddening, the colour excess can be determined from Fig. 3 to be 1·0 mag. mean for the group of three reddened clusters.

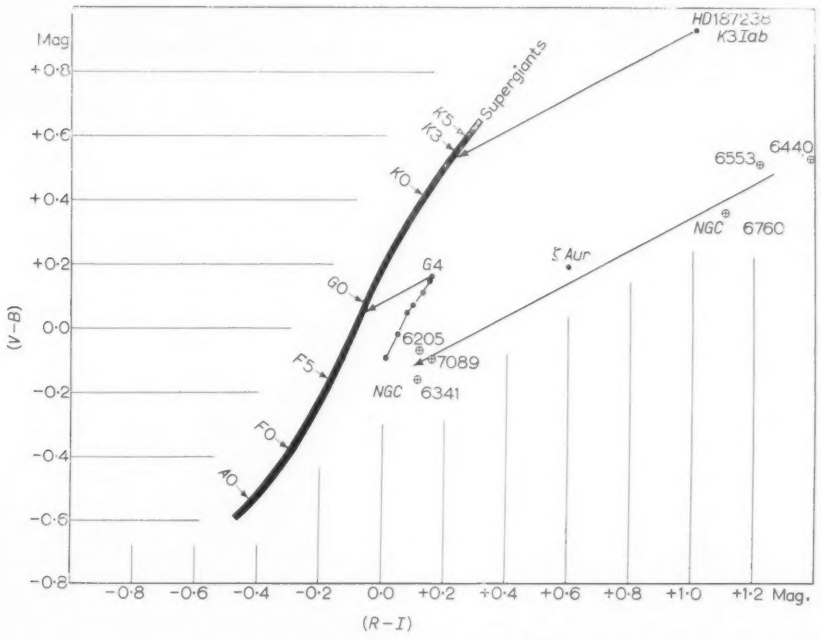


FIG. 3. Supergiant locus from A5 to K5, with plots of three-colour data for a number of interesting objects. Photometric system as in Fig. 2.

When much data for a given class of objects are plotted on a diagram like Fig. 2 or Fig. 3, there is usually evidence of considerable scatter; that is, a star of given spectral type and known luminosity class cannot always be brought down a reddening locus to a satisfactory intersection with the appropriate thermal locus at the given spectral type. Some of this scatter no doubt is experimental error in the photometric and spectrographic data. Of the experimental data, possibly the one known with least precision is the luminosity class, and one might expect to be able by means of photometric methods to measure luminosity among the late type stars with fair precision. Such an accomplishment, however, would depend upon the assumption that multiple colour diagrams would be scatter-free if the experimental data were free of error, and, in any event, the precision of any kind of information obtained by the use of a

diagram like Fig. 3 would naturally be reduced by the presence of intrinsic scatter in the photometric data. In a comparison between two sets of colour difference data believed to be of high accuracy, KRON, GASCOIGNE and WHITE (1957) found an apparent real lack of high correlation of the colour differences, even though the test stars were selected for a high degree of homogeneity in luminosity category.

A possibility of separating the effects of luminosity from those of reddening by means of multiple colour photometry without the help of spectroscopy lies in employing more than three colours. The ultraviolet colour as a component of a three-colour system gives great sensitivity to reddening, but relatively little to luminosity. On the other hand, systems involving the infrared give enhanced sensitivity to luminosity, at least for stars of spectral type later than about G0. Statistically, of course, one can separate luminosity classes by means of three-colour photometry for large masses of evidence without spectrographic data, for stars in regions of the sky where reddening may be expected to be absent or uniform, as has been shown by BECKER and STEINLIN (1956). The general problem for a specific single normal star may be at least partly solved by means of forming three colour-differences from, say, the six-colour photometry. On a three-axis co-ordinate system, the colour differences (U–V), (V–B), and (R–I) were plotted on respectively the Z, Y, and X axes, for a number of representative stars found in the six-colour data. In this three-dimensional diagram, the mean unreddened thermal loci are space curves, and the displacement caused by reddening takes place on a surface of which the thermal locus is an edge. Luminosity differences show themselves among unreddened stars as a preference for plotted points to lie on or between thermal loci of one or another mean luminosity class. The plotted points for stars that have reddening are in or near surfaces, each surface representing a mean luminosity class populated by the data of stars of various degrees of thermal and space reddening. When a three-dimensional plot of the type just described is constructed on a flat surface, it is found to be very complex, and difficult to interpret, especially if the observer happens to be looking endwise down one of the space curves, and therefore sees it in such unfavourable perspective that at least one additional diagram employing a different projection must be constructed. Its complexity prevented a critical interpretation of an experimental plot, but the plot did give encouragement. Indications were that four-colour data properly exhibited may constitute an effective method for classifying single stars over sufficiently wide ranges of reddening, spectral type, and luminosity, to be useful.

6. Colour Temperatures

The temperature of a body of known radiating properties can be measured by comparison with a standard of known temperature by means of a two-colour measurement. When this process is applied to stars, it is found that the properties of stellar radiation are so complex that the determination of a colour temperature is better done by means of a more comprehensive type of photometry. The effects of irregularities of stellar radiation in comparison with black body radiation can be better appraised and averaged if the sampling of radiation is done in many wavelengths. A spectrophotometric approach is probably the best, but it is tedious. Multiple colour photometry can be used effectively, especially if there are many colours and if the inclusive colour base line is long.

STEBBINS and KRON (1957) have recently measured the colour of the Sun by

means of the six-colour photometry, and have at the same time compared the Sun with a standard tungsten lamp (Fig. 4). This general procedure has been used many times for determining colour temperatures of stars (for example, H. KIENLE, 1930), but not photoelectrically with a colour system as suitable as the six-colour one. The comparison between standard lamp and unknown can be interpreted as temperature of the unknown only if the temperature of the standard is known, and if the effective wavelengths of the colour regions are known for both standard and unknown, because the wide spectral regions of multicolour photometry produce shifts in effective wavelength with the colour of the observed object. The temperature of the standard and the effective wavelengths of the colour regions are quantities difficult to determine. Generally, effective wavelengths are found by observing with the photometer

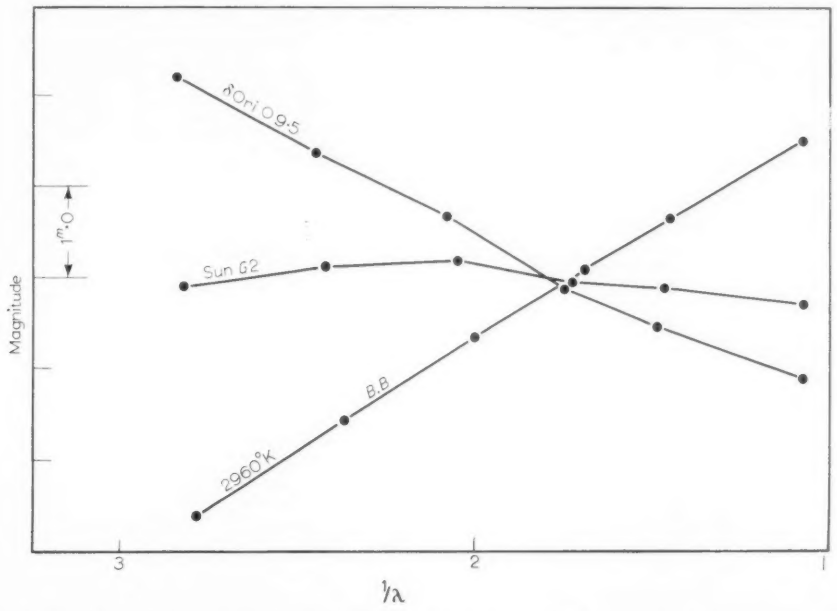


FIG. 4. Plot of six-colour photometric data for a black body, the Sun, and a hot star, showing large range of colour and shifts in the effective inverse-wavelengths of the six colours.

a light source of known relative intensity with wavelength through the filters. The method often used is to pass light from a tungsten filament through a monochromator, and calibrate the emergent light in intensity units by means of a thermopile whose receiver is assumed to be black. If the receiver is not black, no serious error will be made in determining effective wavelength unless the function expressing departure from blackness has within it sharp curvatures or discontinuities, an unlikely possibility in the range of the optical spectrum that is passed by the main transmission band of the atmosphere of the Earth. The measurement is particularly difficult in the ultraviolet colour region, owing to the low intensity of tungsten radiation here; however, good measurements are possible in the six-colour ultraviolet, as well as in the rest of the required wavelength regions. Observations on the now calibrated light from the monochromator are then made in many wavelengths with the stellar photometer, through each filter of the colour system. The thermopile calibration of

the monochromator light makes possible the calculation of sensitivity curves for each colour of the astronomical photometer, as if observations had been made on a light source of uniform intensity with wavelength. Such a fictitious light source is usually called an "equi-energy" source. With these data available, the rest is straightforward, though tedious. For any light source of known or approximately known spectral content, effective wavelength can be calculated for each filter-cell combination by means of a mechanical integration process; refined work should also include the effects of the spectral transmission of the astronomical telescope and of the atmosphere of the Earth when necessary. For the purpose of the effective wavelength calculation, the spectral content of the light from most stars can be determined accurately enough by assuming a temperature based upon the spectral type and luminosity, and assuming black-body radiation.

The temperature of the standard lamp fixes the zero point of the colorimetric method of determining temperatures. The temperatures of most stars, however, are so much higher than those of most practicable terrestrial sources, such as tungsten lamps, that large extrapolations are necessary, which places serious demands upon the accuracy of the standard. For example, if a tungsten standard at, say, 3,000°K is out by 10° (an error of only 0·3%), then the colour temperature of the sun, based on this standard, would be out by 40° for this reason alone, and the temperature of a hot star at 25,000°K would be out by 1,000° or 4%. But the real situation is worse than this. The tungsten is only a secondary standard. The primary standard for high-temperature measurements by radiation methods is the temperature of the melting point of gold, a temperature high enough to be measured by photometric methods, yet low enough to be calibrated with such instruments as thermocouples and gas thermometers without their destruction. The accepted and internationally used value for the gold point is now 1,336°K, but this may be uncertain by as much as 1°, resulting in an uncertainty from this of 20° at the solar temperature, and 500° for a star of 25,000°K. Additional errors are made on the way, however. The local tungsten standard must be compared with a laboratory tungsten standard which has been calibrated with a gold point apparatus. The comparison between the two tungstens can be made with high accuracy (KIENLE, 1936), but the comparison with the gold point is something else. As if this were not enough, there still remains the matter of scale, which is determined by application of the PLANCK radiation law. An additional appreciable uncertainty is present here because of lack of exact knowledge of the radiation constant c_2 that appears in the expression of this law.

Everything taken into consideration, the temperatures assigned to a source of about 3,000°K by different laboratories are likely to differ by 20° to 30°, equivalent to 2,000° to 3,000° at a temperature of 25,000°K, or of the order of 10%. If this seems to be acceptable accuracy, it should be remembered that temperatures are frequently used for calculating total radiation, which goes as T^4.

The field of multiple colour photometry in astronomy contains fascinating vistas for the future. Very little exploration has been made except in three-colour photometry, and even here we have made only a good beginning. The possibilities of the infrared spectral region as a colour in multiple colour photometry are all but unknown, while the future in photometric systems of four and more colours has been only roughly indicated by the accomplishments of the past and present. Some of the work in multiple colour photometry has been ahead of its time, but appreciation is

growing rapidly, and the day may soon be with us when photometry in less than three colours is regarded by all as archaic.

In the preparation of this paper I am indebted to JOEL STEBBINS for helpful criticism and for aid in the preparation of the section on temperature measurement.

REFERENCES

ARP, H. C.	1955	*A.J.*, **60**, 317.
BECKER, W.	1946	*Veröff. Univ. Sternwarte Göttingen*, Nos. 79, 80, 81 and 82.
	1938	*Z. Astrophys.*, **15**, 225.
BECKER, W. and STEINLIN, U.	1956	*Z. Astrophys.*, **39**, 188.
BIDELMAN, W. P.	1957	*Publ. Astron. Soc. Pacific*, **69**, 327.
EGGEN, O. J.	1955	*Astron. J.*, **60**, 131.
GREAVES, W. M. H.	1948	*M.N.*, **108**, 131.
HALL, J. S.	1941	*Ap. J.*, **94**, 71.
	1934	*Ap. J.*, **79**, 145.
HILTNER, W. A.	1956	*Ap. J. Supp.*, **2**, 389.
JOHNSON, H. L. and HARRIS, D. L. III	1954	*Ap. J.*, **120**, 196.
JOHNSON, H. L. and MORGAN, W. W.	1953	*Ap. J.*, **117**, 313.
KIENLE, H.	1936	*Z. Astrophys.*, **12**, 178.
	1930	*Nachr. Ges. Wiss. Göttingen, Fachgruppe* II, Nr. 3, 90.
KRON, G. E., GASCOIGNE, S. C. B. and WHITE, H. S.	1957	*A.J.*, **62**, 205.
KRON, G. E., White, H. S. and GASCOIGNE, S. C. B.	1953	*Ap. J.*, **118**, 502.
ÖPIK, E.	1925	*Publ. Obs. Univ. Tartu*, **26**, part 3.
STEBBINS, J.	1950	*M.N.*, **110**, 421.
STEBBINS, J. and KRON, G. E.	1957	*Ap. J.*, **126**, 126.
	1956	*Ap. J.*, **123**, 440.
STEBBINS, J., KRON, G. E., and SMITH, J. L.	1952	*Ap. J.*, **115**, 292.
STEBBINS, J. and WHITFORD, A. E.	1945	*Ap. J.*, **102**, 318.
	1943	*Ap. J.*, **98**, 20.
TIKHOV, G. A.	1916	*Izvestiia Russk. Abshch. Liub Mirovedeniia*, **5**, 119
WHITFORD, A. E.	1948	*Ap. J.*, **107**, 102.
WOOLLEY, R. v. d. R., GASCOIGNE, S. C. B. and DE VAUCOULEURS, A.	1955	*M.N.*, **114**, 490.
YÜ, C. S.	1926	*Lick Obs. Bull.*, **12**, 104.

BIBLIOGRAPHY ON SIX-COLOUR PHOTOMETRY

STEBBINS, J.	1955	*Publ. Astron. Soc. Pacific*, **65**, 118.
	1950	*M.N.*, **110**, 416.
	1946	*Ap. J.*, **103**, 108.
	1945	*Ap. J.*, **101**, 47.
STEBBINS, J. and KRON, G. E.	1957	*Ap. J.*, **126**, 126.
	1956	*Ap. J.*, **123**, 440.
	1954	*Ap. J.*, **120**, 189.
STEBBINS, J., KRON, G. E. and SMITH, J. L.	1952	*Ap. J.*, **115**, 292.
STEBBINS, J. and WHITFORD, A. E.	1948	*Ap. J.*, **108**, 413.
	1947	*Ap. J.*, **106**, 235.
	1945	*Ap. J.*, **102**, 318.
	1943	*Ap. J.*, **98**, 20.
WHITFORD, A. E.	1948	*Ap. J.*, **107**, 102.

Experiences in Photometry with the Cambridge Schmidt Telescope

A. N. Argue

The Observatories, Cambridge, England.

Summary

The purpose of this paper is to describe a preliminary investigation of the performance of the Cambridge Schmidt telescope in photometry, and to relate it to the general subject of photographic photometry. Emphasis is placed on the dependence on measuring procedure, particularly towards the photographic threshold. The results suggest that the field errors are low within a diameter of at least three degrees, and that magnitude transfers can be performed, provided that the exposures are of a few minutes' duration only. Mention is made of the use of the EDSAC electronic computer, for reducing the large number of measurements.

1. Introduction

Twenty-five years have elapsed since the introduction of the Schmidt optical system. During this period the most extensive applications of Schmidt telescopes have been to pictorial surveys in various spectral regions and to objective-prism surveys. Some of the latter have been carried out quantitatively, in particular for the determination of spectrophotometric parallaxes. In general, however, the application to quantitative studies has been relatively small. Their use in astrometry has not progressed beyond three or four trials. Photometric programmes have been more common; nevertheless, they still occupy only a meagre proportion of the Schmidt telescope-time available.

At first sight the Schmidt telescope appears ideally suitable for in-focus photographic photometry. It provides high image-quality over a wide field, 5° or greater, so that a single exposure should give images down to a faint limiting magnitude with small field errors. In classical systems with aperture-ratios around f/4 this performance is sufficiently maintained over wavelengths extending from ultraviolet to infrared to fulfil the requirements of modern three-colour photometric surveys; see, for example, Linfoot 1955.

In spite of these considerations it has turned out that certain Schmidt telescopes are not, in practice, suitable for general photometry. The reason given is the small size of the image, making the measured magnitude critically sensitive to certain disturbances, such as errors in adjustment, focus changes and differential flexure. The classical Schmidt requires the photographic emulsion to be bent to the Petzval surface. It is not possible to do this exactly by pressing a plate by its edge onto a spherical surface, and the focus errors are likely to vary from plate to plate on account of elastic variations. Again, as different portions of the mirror are used in different parts of the field, irregularities in figure, such as those caused by flexure or thermal distortion (Couder, 1955), introduce irregular errors. An uneven reflection coefficient of the mirror may be troublesome. Multiple exposures are affected by changes in seeing and guiding. For a Schmidt of short focal-length, where the diameter of the threshold image is determined by photographic emulsion-spread, a

lower signal-to-noise ratio is obtained than for an instrument giving larger images, in which the effective number of grains is larger (Haffner, 1956).

There arises then the question of whether the classical Schmidt is fundamentally unsuitable for photometry. Should it turn out that no mechanical mounting can be devised to give adequate stability under all likely conditions of orientation and temperature, and that no suitably matched photographic emulsion can be procured readily, then the system must be deemed unsuitable. Already proposals have been made, and indeed carried out, to modify the system by under-correcting even to the extend of removing the corrector plate altogether. It seems, however, that the situation has not yet been fully explored, and that the time has not yet come to answer the question of fundamental suitability.

Further problems are raised by the enormous amount of information potentially available from Schmidt photographs extending to faint magnitudes over wide fields. For instance, there is no difficulty in obtaining 10^5 images on a single photograph with an exposure time of less than half an hour. However, traditional methods of measurement and reduction permit only a small fraction of these to be handled, if one is to keep pace with a fairly regular supply of plates from the telescope. Recent improvements in irisphotometer design have enabled a substantial increase in measuring speed to be obtained. For instance the machine developed by W. Becker and his associates (Becker and Biber, 1956), which the author has had the privilege of using in Basel, allows a skilled operator to identify and measure for density some 300 stars per hour. This is probably about the limiting speed that can be obtained by hand. Although slow compared with the output from the telescope, nonetheless it introduces the problem of processing the large amount of data obtained. In Cambridge we have been fortunate in having the EDSAC 1 and 2 electronic computers placed at our disposal, offering a way of clearing reductions easily.

It was to examine all these problems that some trial photometric programmes have been carried out with the Cambridge Schmidt telescope under the supervision of Prof. Redman. It is not proposed to describe these in detail here. Some of the earlier work was done jointly with my colleague Dr. Dewhirst, who had also undertaken a large part of the initial adjustment of the new telescope.

The purpose of the present article is to describe some preliminary results of general application to Schmidt photometry in the hope of contributing to a review of the subject, and to illustrate the use of the electronic digital computer in reducing even a small fraction of the data available on the plates.

Because of the nature of the procedure, the result of a magnitude determination by in-focus photographic photometry depends not only on the structure of the image formed by the telescope, but also on the manner by which the image is subsequently measured. In consequence, the telescope cannot be considered in isolation; it must be examined in relation to the measuring process, and a discussion of the measuring technique (carried out in this instance by a Sartorius Irisphotometer) must therefore be included.

2. The Cambridge Schmidt Telescope

This is a 17/24/64 inch classical Schmidt, with a field of 5° of which 3°3 are unvignetted.

The mirror is of Pilkington low-expansion glass. The lateral thrust is supported

by six counterbalance units, acting on the side of the mirror, and the lateral position is determined by two stops and a spring plunger. The back of the mirror rests on eighteen fluon discs mounted kinematically in the normal manner.

The corrector is of selected white plate-glass, of meniscus form, and figured on its back surface. The corrector ghosts are of very low intensity because of this meniscus form and of the fact that both surfaces are bloomed.

Circular photographic plates are used and pressed against the spherical back of the plateholder. Temperature compensation of focus is provided by four brass-invar links from the plateholder supports to the mirror cell. This compensation works well for slow temperature changes; there appears to be no effective method of dealing with rapid variations in temperature.

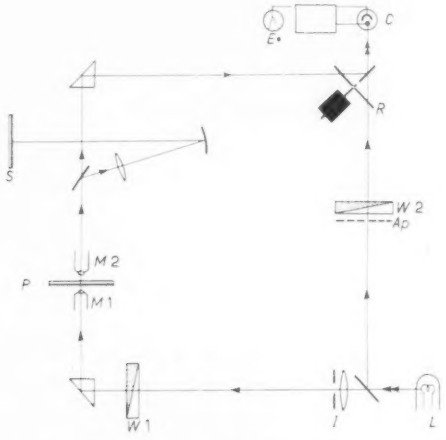


Fig. 1. Schematic diagram of Sartorius Irisphotometer.

The drive is by a 47 c/sec synchronous motor, fed by a resistance-capacity oscillator and amplifier. The frequency can be adjusted through ± 1 per cent. The worm, and drive wheel are continuously lubricated. The guiding telescope is a 6/90 inch refractor. No provision exists for guiding on an image formed by the main telescope. Recent experience has indicated that such an arrangement would be desirable. In exposures exceeding one hour there is an elongation of the images by one or two seconds of arc, arising, it is thought, from motion of the plateholder support. This elongation generally varies over the field so that a magnitude-dependent field-error would be expected; no investigation of this has yet been made.

3. The Sartorius Irisphotometer

The image densities were measured with the Sartorius Irisphotometer borrowed from the Royal Greenwich Observatory (by courtesy of H.M. Astronomer Royal). A description of this instrument has been published by H. Haffner (1953). The lay-out is indicated by Fig. 1. An iris diaphragm I, illuminated by a lamp L is imaged greatly reduced on the photographic plate P by the microscope objective M1. The density wedge W1 is turned right out of the beam for normal measurement. Some uses for it are described below. The beam then passes through a second microscope objective M2 and up to the photocell C. A comparison beam also passes from

L to C via the aperture Ap and the density wedge W2. The rotating shutter R passes the beams in rapid succession, and an electronic indicator E shows when both beams are equal in intensity.

When a star image is centred in the iris image on P, a proportion of light is absorbed and the iris diaphragm must be opened to restore equality of intensity of the two beams. An arbitrary scale attached to the iris movement thus gives a measure of the photographic effect of the star. This measure is independent of the brightness of the lamp L in so far as the balance of the beams is also independent. However, during runs extending for several hours, drifts in the zero-point of a few hundredths of a magnitude were observed. The effect of switching off overnight occasionally introduced scale and zero-point-errors exceeding 0·05 mag. These are thought to be temperature effects. In order to use a single calibration curve through one run, it is necessary to maintain a constant reduction factor for the iris image formed by M1. This can be done by leaving M1 untouched, and by adjusting the height of P and M2 to maintain the iris and the photographic grain in focus on the screen S. It was found necessary to repeat this at intervals of 1 to 2 hours, during which time a displacement of the structure supporting M2 had occurred. By maintaining this adjustment, the overall drifts during a day's run rarely exceeded 0·03 mag.

The effect of switching off was thought to arise mainly from variation in the position of the lamp filament. When the effect exceeded 0·05 mag, a separate calibration was made. In all cases the machine was switched off overnight, and switched on by a time-switch three hours before measuring.

When the measurements were read out to a secretary, 1,000 stars required about eight measuring hours. This is an appreciably slower rate than is possible with the Basel photometers already referred to. The difference lies mainly in the balance indication which is a little diffuse on the Sartorius, depending on the image quality, so that three or four repetitions are necessary before the iris reading can be accepted with confidence. Carelessness in this respect can easily lead to errors exceeding 0·1 mag. When a sample of the same plates was measured on a Basel photometer, however, the iris settings could be accepted with much fewer repetitions.

4. Reduction on EDSAC

The readings were written down, corrected for drift, and then punched on paper tape, together with the magnitudes and colour indices of the calibration stars. Punching and tape preparation required about four man-hours per 1000 stars. It would of course be more efficient to punch as the readings come from the photometer, or even to modify the photometer to punch directly, applying any drift correction as part of the computation programme.

The relationship between iris-reading and magnitude is expressed by a polynomial whose coefficients are determined by least squares. The degree m is set up on the EDSAC manual register during the programme input routine. Values of m as low as 3 sometimes give a sufficiently close fit, but usually m is set around 8. Higher values such as 12 have been found to make only negligible difference to the results, while they slow down the rate of working very considerably. Examples of calibration curves are given in Fig. 11.

The first solution is sensitive to gross errors in data, such as punching errors, mistaken identities, etc. The machine tests the residuals; those which are greater

than a certain prescribed amount are temporary rejected in forming another solution, and this process is continued until a steady state is reached.

The data tape is run into the machine twice. In the first run the standard stars are picked out to form the equations of condition. So far all magnitudes have been derived on the local R.G.U.-system (Becker and Steinlin, 1956). The calibration magnitudes, however, have always been on the U.B.V.-System (Johnson and Morgan, 1953), so that in the case of the R and G plates it has been necessary to apply preliminary colour corrections of the form

$$R - V = f_R\,(V, B-V)$$

for R, and similarly for G. (The luminosity may be expressed by V because in all cases the calibration stars are members of some cluster.) The standard magnitude V is corrected to R (or G) by the machine before insertion into the equations of condition. On reaching the end of the tape the machine punches out the co-ordinates of about 25 points on the calibration curve for hand-checking. On running in the data tape a second time, the identity and magnitude derived for every star are punched out. For 1,000 stars the machine-time required is about 5 minutes, most of which is punching-time. A sample of the process is given in Fig. 2.

Another programme has been devised to average individual magnitudes for each star as recorded on up to 63 tapes. At present the storage capacity of EDSAC 2 (1,024 registers of 40 binary digits) limits this programme to 946 different identities; any extra identities are punched out.

5. Measurement with Irisphotometer

Let $I(r,\theta)$ be the intensity of illumination of the iris image, and $t(r,\theta)$ the transmission coefficient of the photographic image at the same point. The total intensity L transmitted by the photographic image is given by:

$$L = \int\int_A Itr \cdot \mathrm{d}r \cdot \mathrm{d}\theta \qquad \ldots\ (1)$$

As an approximation, I is constant within the iris image and is zero outside. One may then regard the range of integration in equation (1) to be over the iris image. To make a magnitude measurement, this range is adjusted to make L equal to a constant L_0, the intensity of the comparison beam. The arbitrary scale attached to the iris movement gives a measure A of this range of integration. The magnitude derived from A contains, along with others, the following sources of error:

(1) the experimental error in balancing L against L_0;
(2) error in centring the star image in the iris;
(3) random fluctuations of $t(r,\theta)$, that is, grain noise;
(4) variations in background fog;
(5) systematic differences in image structure, leading to differences in $t(r,\theta)$ among images of the same magnitude in the system.

(a) *Balancing Error*

This error is a minimum when $(1/L)(\mathrm{d}L/\mathrm{d}A)$ is a maximum. On the Sartorius instrument a convenient method is available for determining the corresponding A for a

given image. One centres the image accurately, and for a number of iris settings A one finds the corresponding $W1$ wedge settings to balance, the comparison-beam intensity being constant; $W1$ is shown in Fig. 1. A plot of A against $W1$ is given in Fig. 3. Since the density of this wedge is a linear function of its reading, $W1$, it follows that

$$\frac{1}{L}\frac{\mathrm{d}L}{\mathrm{d}A} = -k\frac{\mathrm{d}W1}{\mathrm{d}A}$$

where $-k$ is the wedge constant. The balancing error is therefore least where the slope in Fig. 3 is numerically greatest. The star used for Fig 3 was of the 10th magnitude, which was five magnitudes above the limit of the plate. Where the slope is greatest, the iris has cut into the outlying grains of the image and is in a region where the density is low (about 0·2 above sky-background) but changing perceptibly.

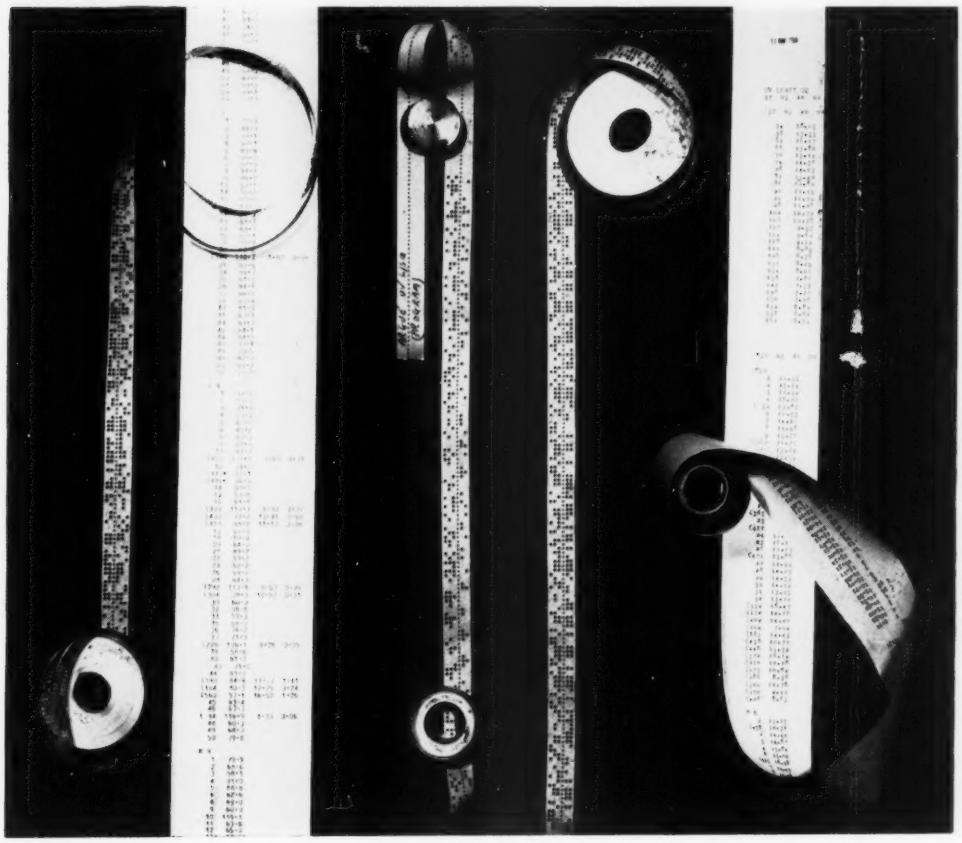


Fig. 2. EDSAC 2: punched tapes.

The tape at the extreme left contains the data from the photometer and the standard magnitudes. Beside it is its print-out. The centre-tape contains the programme. On running it into EDSAC, followed by the data-tape, the calibration curve is punched out onto the right-hand tape. This is followed by the magnitudes of all stars, on running in the data-tape a second time. On the extreme right is a print-out of the output-tape; the calibration points are in the top block, the identities and magnitudes below.

Fig. 3 shows that the optimum iris setting for balance is not critical. Because of this, a single adjustment of the comparison beam is sufficient over a range of about ten magnitudes.

(b) *Centring Error*

The magnitude error is approximately proportional to the square of the off-centre distance. It is roughtly independent of magnitude for a given adjustment of the comparison beam, but increases when the latter is changed, so that the iris closes more tightly around the image. For the Cambridge Schmidt the usual adjustment gives an error of 0·03 magnitudes for a centring error of 4μ. On the screen of the Sartorius this corresponds to about 0·2 mm; thus careful centring is required.

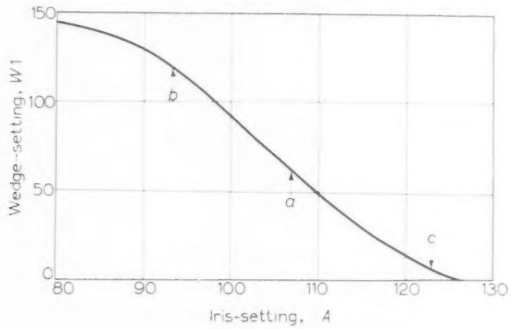


Fig. 3. Sartorius wedge-setting $W1$ plotted against Iris-setting A, for constant transmitted intensity L, according to equation (1). Star 10th magnitude, plate-limit 15th magnitude.

(c) *Consistency of Measurement*

The internal consistency of measurement of a red plate is demonstrated in Fig. 4 by *a* and *b*. In *a*, for instance, are plotted for each image the differences in measured magnitudes obtained in two measuring runs. The iris was approximately at optimum adjustment for balance at 10 mag, corresponding to the point *a* in Fig. 3. In Fig. 4 *b* the differences are plotted for two runs with a smaller iris. On a 10 mag.-image this iris cut into the region where the photographic density was 0·7 above background, and corresponded to *b* in Fig. 3.

The consistency of measurement is noticeably better in *a* than in *b*, except in the region beyond 14 mag, where for *a* the calibration curve is considerably flatter. No extraordinary care was taken over these measurements. They were made at the usual rate, and should be representative of the measuring accuracy ordinarily achieved in survey work. From the point of view of measurement it is concluded that the choice of iris adjustment between these two values is not critical, unless one is concerned particularly with the images near the threshold of the plate. For these, Fig. 4 shows the small iris to give greater measuring consistency, but evidence presented later in connection with granularity and focus-change suggests that the large iris is in fact preferable.

From Fig. 4 *a* and *b* the internal probable error of a single measurement is less than 0·01 magnitude.

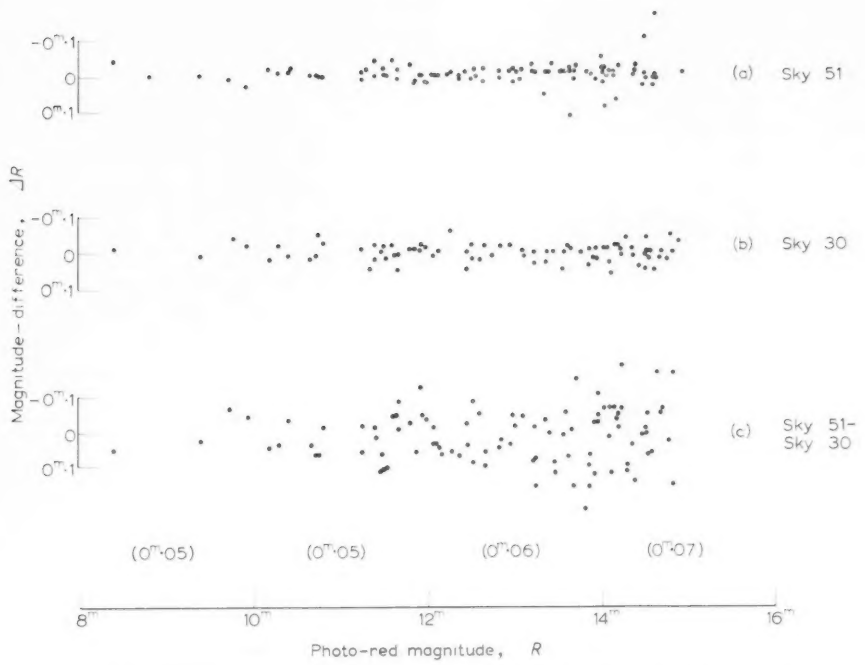


FIG. 4. Internal consistency of measurements with the Sartorius Irisphotometer. Plate limit 15th magnitude.
(*a*) Large iris (SKY = 51); (*b*) small iris (SKY = 30); (*c*) Comparisons of single measurements in *a* and *b*.

(d) *Granularity*

On the other hand when one magnitude measurement of *a* (Fig. 4) is subtracted from one of *b*, the plot *c* is obtained. This plot therefore compares the magnitudes obtained by measuring over two regions of the image, that for *b* being included in *a*. The differences in *c* show no correlation with colour index, such as one may perhaps expect from wavelength-dependent scattering in the emulsion. They appear to arise entirely from grain in the image. The numbers in parenthesis are the average difference, without regard to sign, in each 2-magnitude interval.

By means of Fig. 3 it is possible to determine the transmission distribution $t(r)$ in equation (1), ignoring variation with θ. The mean square fluctuation $\overline{(\delta L)^2}$ over the region within the iris was then calculated; it is plotted in Fig. 5. In this calculation it was assumed that the density fluctuation over an element of area a is given by a Gaussian error distribution with root-mean-square deviation $\sqrt{(G^2/2a)}$, where G is SELWYN's *graininess constant* (MEES, 1942). Values of G were obtained from microphotometer tracings of a contact print of a step wedge on the same emulsion (Kodak 0a-E).

δL was converted into r.m.s. magnitude-fluctuation and plotted in Fig. 6. Great weight should not be given to this because the graininess measurements were very rough, being considerably influenced by defects in the wedge and emulsion, and because no account was taken of variation of G with a. Nonetheless, Fig. 6 may be of value in indicating the dependence of δm on the iris reading A. It suggests for this

particular image, 5 mag above the plate limit, an uncritical minimum grain-noise in the region of a, the point corresponding to optimum adjustment for balance. It also indicates increased noise for a large iris radius, due to the sky-background grains.

These noise calculations now offer an explanation of Fig. 4*c*. At 10 mag, the magnitude fluctuations are according to Fig. 6 about equal for the two iris adjustments

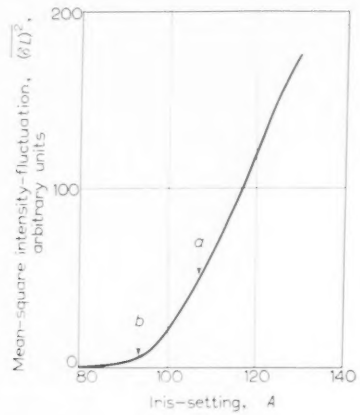


FIG. 5. Plot of the mean-square fluctuation in intensity L transmitted by the star-image against Iris-setting A. Star 10^m, plate-limit 15^m.

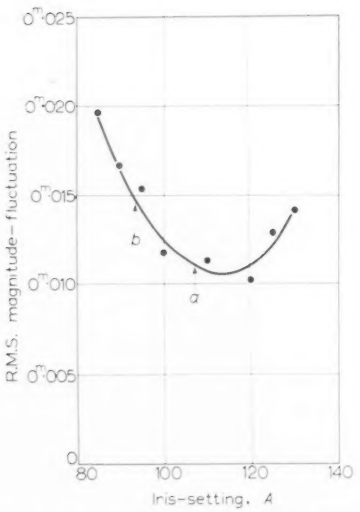


FIG. 6. Calculated r.m.s. magnitude-fluctuation against Iris-setting A on balance. Star 10^m, plate-limit 15^m.

corresponding to a and b (Figs. 4 and 6). Although these fluctuations are not independent because the whole region of the image measured in b is included in a, they may be regarded as approximately so. This follows from Fig. 5 which indicates the noise power for b to be one tenth that for a. Accordingly the r.m.s.-difference in

a magnitude-interval in Fig. 4*c* is approximately $\sqrt{2}$ times the r.m.s.-error of one magnitude measurement. This last quantity is given in Table 1.

Table 1

Magnitude	8–10	10–12	12–14	14–15
r.m.s. error	$0^m.03$	$0^m.04$	$0^m.04$	$0^m.05$

R.m.s. error of a magnitude measurement due to granularity and error of measurement.

Cambridge Schmidt; Kodak 0a-E; Chance OR1 filter; Exposure 15 minutes; Sartorius Irisphotometer.

The values in Table 1 indicate the noise calculated for 10 mag, (Fig. 6), to be low by a factor of about three. This is, however, in good accord with the internal deviations experienced with the R magnitudes in a three-colour survey of the Praesepe region (as yet unpublished). From this it is concluded that the magnitude error, at least over one degree square near the centre of the plate, is mainly due to granularity.

A calculation may also be made of the grain-noise near the threshold, where the image consists of a uniform-cluster of grains of a diameter of about 30μ. When optimized for balance five magnitudes brighter, the iris just encloses this threshold cluster. The r.m.s.-fluctuation δm is then given approximately by (FELLGETT, 1958):

$$\delta m = \frac{\Delta D_{10}}{0{\cdot}49\,\gamma}\sqrt{\frac{4}{\pi}} \qquad \ldots . (2)$$

where ΔD_{10} is the r.m.s. density-fluctuation for a circular scanning element of diameter 30μ. Taking $\Delta D_{10} = 0{\cdot}034$, and gamma unity, gives $\delta m = 0{\cdot}08$ mag, a result open to the same criticisms as those made against Fig. 6.

(e) *Background Density*

Local increases in background density can be caused by scattering and reflection at the optical surfaces. In addition to the familiar corrector-plate ghost, each very bright star on our R.G.U.-plates is situated in a disc displaced towards the centre of the plate. This disc is a faint out-of-focus image of the star formed by one reflection at each surface of the flat filter, placed immediately before the emulsion. Discs formed by further reflections do not show up, but may affect the photometry of nearby stars. Of course, these discs only appear for very bright stars, and they could be reduced in intensity by blooming.

It has been mentioned how the corrector has been arranged to reduce the intensity of its ghosts, so that normally these do not show up. When they do, they offer a convenient method of examining the effect of background fog on the magnitude because the position of the ghost in the field may be controlled. Table 2 gives measurements of four stars which coincide with one of these ghosts on one U-plate, so that the background density had a local increase of 0·0086 (log., units). Column 2 gives the mean U from other images not affected by ghosts. Column 3 gives U′, obtained by taking no account of the increased background density. (U′—U) in the next column shows the pronounced effect of the ghost.

The comparison beam was then altered so that the iris reading on the ghost-background was equal to the value it had on the undisturbed background in other regions, and U″ was obtained. Although the observations are meagre, the residuals (U″ − U) suggest at least partial compensation by this process. The wedge W1 (Fig. 1) is helpful in facilitating adjustment on a large area of background.

This process of compensation is analogous to the procedure described by Ross (1936) for the Schilt photometer, in case where the fog increases the density of star-image and background by the same amount. For images of low density one may

Table 2

Effect of altering the comparison beam to compensate for background density.

Star No.	Mean U no ghost	U′ in ghost	U′—U	U″ compensated	U″—U
	m	m	m	m	m
3	14·79	14·09	−0·70	14·79	0·00
4	15·49	15·03	−0·46	15·44	−0·05
5	15·62	15·65	+0·03	15·85	+0·23
9	14·23	13·72	−0·51	13·88	−0·35
			−0·41		−0·04

expect this condition to be satisfied approximately, on the basis of the density-exposure curves given in Ross' paper.

(f) *Image Structure*

In a general way equation (1) shows that the measured magnitude is sensitive to changes in image structure, that is, to systematic changes in $t(r,\theta)$ among stars of equal apparent magnitude. This may be illustrated, for example, by the effect of small focus changes in the telescope. Fig. 7 shows this for five stars ranging from 8 to 14 mag, the plate limit being 15·2 mag. The measurement was made with three adjustments of the Sartorius: *a* (centre column) and *b* (left), corresponding to the respective settings of Fig. 3, discussed above; and *c* to a large iris, whose sky-area was 2·5 times that of a faint in-focus image. These faint in-focus images were just cut into for *a*; for *b*, however, they were well cut into, two thirds of the area being cut away.

Although the curves are distorted by photographic grain, they show the general pattern obtained from a large number of these focus sequences. For stars brighter than 10 mag, which is 5 mag above the plate limit, the magnitude effect of focus change is independent of both brightness and iris-adjustment. At 14 mag, the effect of a small focus change (less than 0·1 mm) is similar to that for a bright star for setting *a*, but of opposite sign for *b*. The setting *c* seems to give little effect at 14 mag; however it is not recommended to use this adjustment, because the iris is less sensitive to balance (*c*, Fig. 3) and the slope of the calibration curve undesirably low (see the top pair of curves in Fig. 11).

Taking the iris-adjustment *a* in Fig. 7 as typical, the magnitude error Δm is approximately proportional to $(\Delta f)^2$ for Δf less than 0·1 mm, in confirmation of a result obtained by Haffner (1956) with the Bergedorf Schmidt. From Fig. 7 *a* one obtains $\Delta m = 0{\cdot}018$ mag, when $\Delta f = 0{\cdot}025$ mm. This situation refers to a blue-sensitive plate and minus U.V.-filter, i.e. to Pg-photometry. The corresponding

values for R.G.U.-photometry are included in Table 3. No explanation is offered for the Pg error being the greatest. It may be because the Pg waveband is the widest, or because the minus U.V.-filter is more than twice the thickness of the others. The corresponding figure given by HAFFNER (1956) for the f/3 Bergedorf Schmidt is $\Delta m_{Pg} = 0{\cdot}01$ mag.

Table 3

Magnitude error for bright stars, introduced by a focus error of 0·025 mm, on Cambridge Schmidt (f/3·8)

Pg: Ilford Zenith; Wratten 2B (−UV); G: Ilford Zenith; Chance OY8;
R: Kodak 0a-E; Chance OR1; U: Ilford Zenith; Schott UG2.

Pg	R	G	U
$-0^m{\cdot}018$	$-0^m{\cdot}008$	$-0^m{\cdot}006$	$-0^m{\cdot}013$

Fig. 7 and Table 3 indicate the precision to which the figure of the emulsion surface, the focus and the tilt of the plateholder must be adjusted and maintained. An approximate adjustment for tilt can be achieved quite rapidly using the autocollimation method described by DEWHIRST and YATES (1954), and for this a convex mirror has been provided by GRUBB PARSONS. However it has been found necessary to determine the final correction on the sky by focus sequences; but the autocollimation apparatus is convenient for applying this small remaining correction. It has also been valuable in studying flexure of the plateholder relative to the main mirror. The graph in Fig. 8 shows by how much the centre of curvature C′ of the emulsion surface lies above that of the mirror (C) when the tube is horizontal in the North–South direction, given that they are coincident when pointing vertically upwards. In the East–West direction the flexure is about 40 per cent smaller. Negative focus settings correspond to positions of the plateholder closer to the mirror; hence to greater extension of the supporting arm, and the increased flexure which is further increased by the weight of the filter (400 g). The normal working range of focus is $\pm 1{\cdot}3$ mm, and then with the 2B filter CC′ = 0·2 mm. This corresponds to a focus error at the edge of the field of 0·01 mm, so that $\Delta m_{Pg} = 0{\cdot}003$ mag. It is concluded that the effect of changing tilt of the plateholder through flexure is negligible.

It is known that there is a flexure between the main mirror and the corrector, since if the centre of the latter coincides with C (Fig. 8) at the zenith, it will be 0·4 mm below it with the tube horizontal. If one regards this as a displacement of the mirror, one would expect it to combine with the sag of the plateholder and the flexure of the guiding telescope to produce elongated images during a long exposure. Elongations not exceeding three seconds of arc have indeed been found on some of the longer exposures (of the order of one hour). These elongations vary in amplitude and direction over the plate, suggesting a rotation of the plateholder superimposed on the lateral sag. They are not the same on photographs of the same field taken under apparently identical conditions. Atmospheric refraction probably contributes at most 0″·4 elongation. How the remainder originates is not entirely clear.

6. FIELD ERRORS

Arising from the manner in which magnitudes are measured photographically, one would expect the field error to depend on the combined optical aberration, guiding,

seeing and emulsion image-spreads, on the total light intensity contributing to the image (including the effects of any vignetting and uneven reflectivity of the mirror), and finally on the adjustment of the irisphotometer. The aberration image-spread will depend, *inter alia*, on the position of the emulsion and on the thermal state of the mirror (this last effect is not likely to be important in photometry with the Cambridge instrument). The focus effect will depend on local irregularities of the figure to which the photographic emulsion has been deformed, which in turn will depend on the elastic properties of the glass.

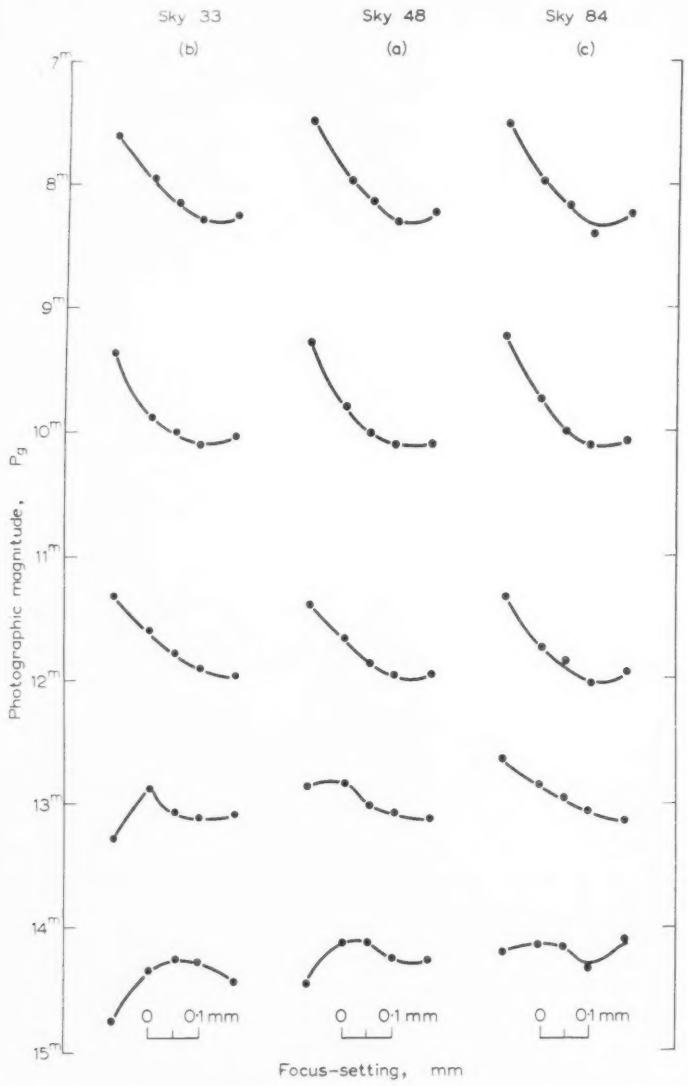


FIG. 7. Dependence of measured photographic magnitude (Pg) on focus-setting for three adjustments of the Irisphotometer.

Cambridge f/3·8 Schmidt; Ilford Zenith plate; Wratten 2B; exposures 2 minutes each.

The concept "field error" may therefore be applied in two ways. In the first, it refers to the variation of measured magnitude on a single plate, due to any instrumental cause whatever, including such random effects as variation in emulsion sensitivity and irregularity in figure not reproduced from plate to plate. From this point of view each plate must be calibrated separately.

Secondly, "field error" may refer only to the average obtained from a large number of plates. In this sense, which is the one usually implied by the term and which is the more convenient in photometric survey, it is a description of the systematic properties of the photometry under the given conditions of observation and measurement.

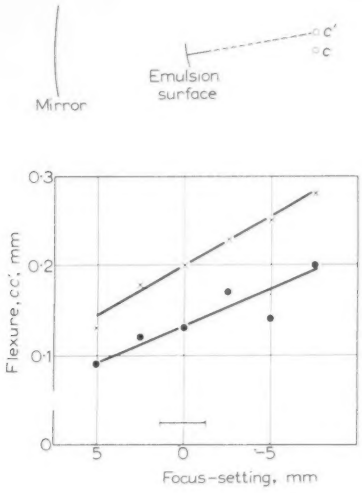


Fig. 8. Variation of flexure CC′ with focus-setting.

CC′ is the distance of C′ (centre of curvature of emulsion) above C (centre of curvature of mirror) with the telescope horizontal in the North–South direction. At that zenith CC′ is zero.
× = with Wratten 2B filter; ● = No filter.

Because of the sensitivity of the Schmidt image to small disturbances, there arises the doubt whether practical systems are sufficiently stable to allow the use of this second concept. For a given telescope this can only be decided by experience. The two field-error determinations (in the first sense), to be discussed, suggest the second concept of field error to be applicable in our particular case, at least when the exposure time is short.

Values obtained for two pairs of exposures on overlapping fields on one U-plate are plotted in Fig. 9 as a function of distance from the centre, with division into four quadrants. Exposure times were 3 minutes each. The magnitudes range from U = 3 to 11 mag, but 90 per cent of the stars are fainter than 7·0 mag. Individual values are plotted in Fig. 9 *a*, and means within radius intervals of 5 mm in Fig. 9 *b*. The same quantities are plotted in Fig. 10 for a G-plate.

For both plates there is the suggestion of an increase in measured brightness in the outer region of the NW-quadrant. The error in focus in this region is certainly not expected to have exceeded 0·05 mm so that, from Table 3, one would expect the maximum effect to be −0·05 mag in U, and −0·03 mag in G.

Taking the unvignetted region (diameter 3°3 of all four quadrants together, we obtain for the average deviations of the individual values (plotted in *a*, Figs. 9 and 10) from the mean, the values 0·032 mag in U, and 0·027 mag in G. These are of the order of the scatter due to photographic grain and measuring error; from this we conclude that the field error is small at every point in the unvignetted region.

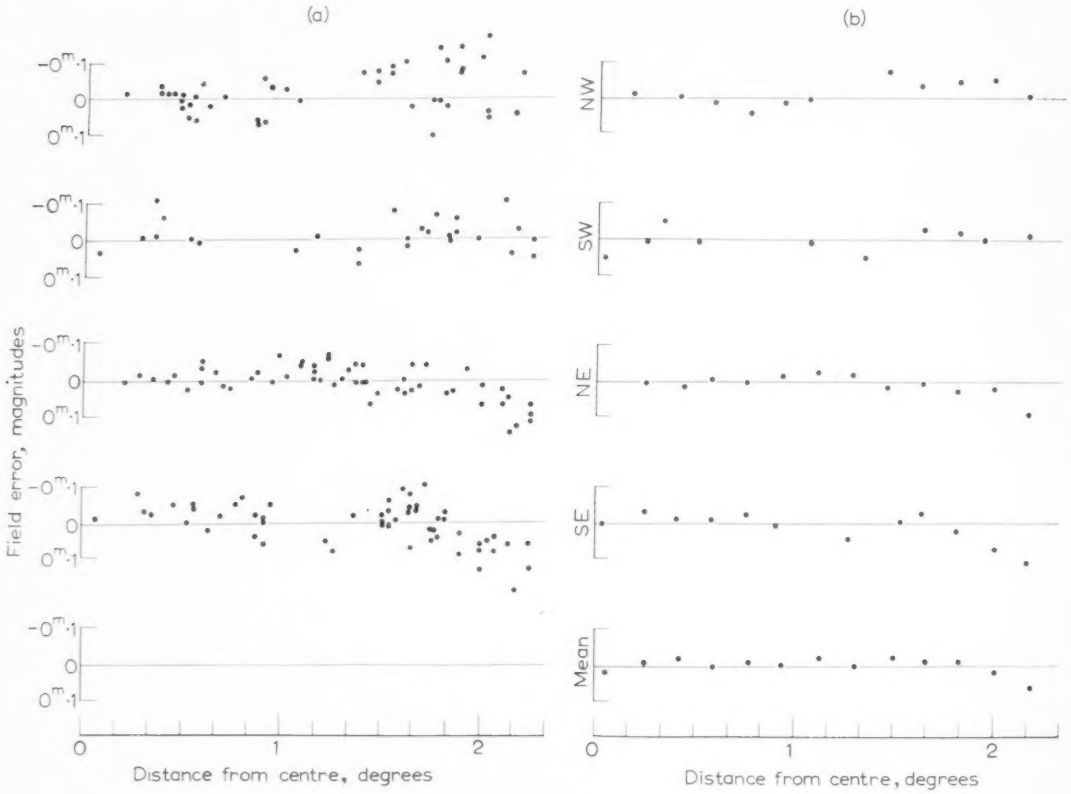


Fig. 9. U field-error of Cambridge Schmidt plotted against distance from centre, with division into four quadrants. Exposures 3 minutes.
(*a*) individual values; (*b*) means.

7. Magnitude Transfer

The calibration curves for a pair of exposures for the same field were obtained with four adjustments of the irisphotometer, and are shown in Fig. 11. They show the familiar effect of changes in seeing and guiding between the two exposures. The differences in magnitude are about the same among the bright stars for all four irisphotometer adjustments. For the faint stars, however, they are small for the large iris (SKY 84) and increase as the iris diameter is reduced.

The points of intersection of these curves correspond to the "optimum density" described by Weaver (1947) for the case of measurement with a Schilt-type photometer. So far no attempt has been made to use the reduction procedure described in his paper. His assumption, that the transformation of one curve to the other

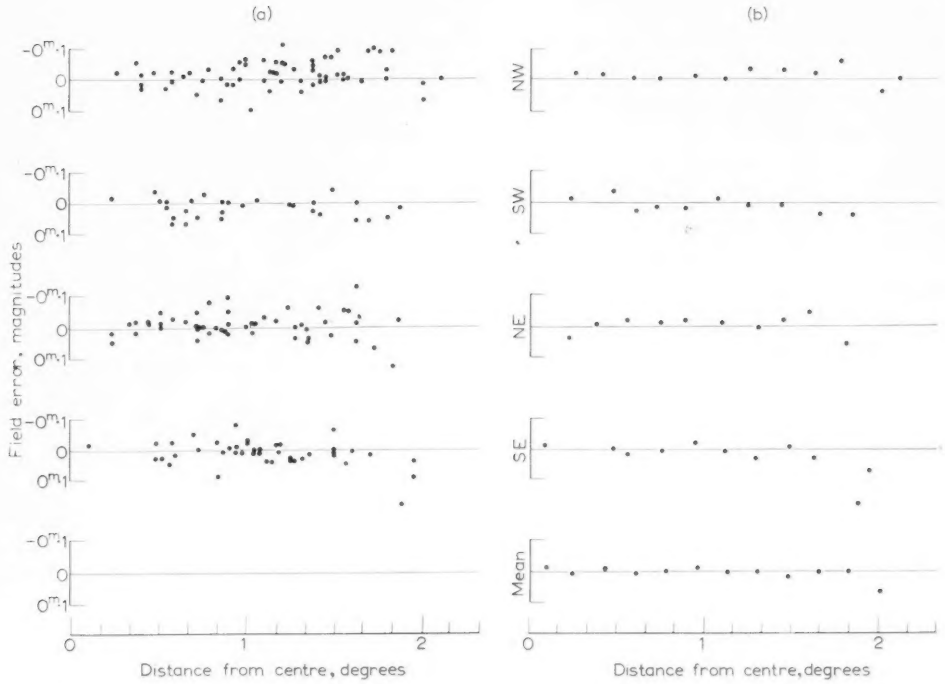


Fig. 10. G field-error of Cambridge Schmidt plotted against distance from centre, with division into four quadrants. Exposure 3 minutes.
(*a*) individual values; (*b*) means.

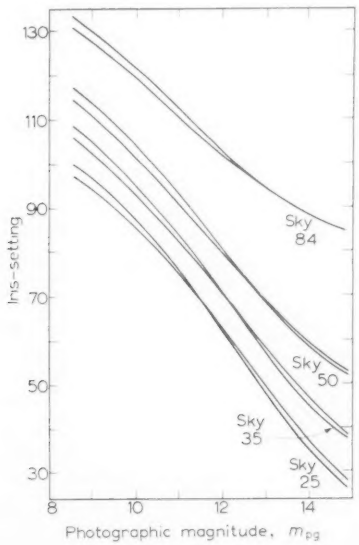


Fig. 11. Irisphotometer calibration-curves for two exposures on one plate and four settings of comparison beam.
Exposures 2 minutes.

takes place by rotation about the point of optimum density, may be inadequate in the present situation; but the curves do suggest that the transfer errors may be reduced over a considerable range of brightness by measuring with two or three different adjustments. However, the scatter obtained for faint stars with large iris (SKY 84) is large.

The small iris-adjustment (SKY 25) is sometimes urged on the ground that here the calibration curve is steepest and most nearly linear. It has already been noted, however, that a small iris, cutting heavily into the faint images, gives rise to a reduction in signal-to-noise ratio, to an increased dependence on centring in the iris, and to a greater sensitivity to focus error in the telescope. A linear calibration is of no advantage when the reduction is performed either graphically or by electronic computer.

As a compromise, the adjustment used for measuring the Cambridge plates has been in the region of SKY 50.

Fig. 11 is typical of results obtained in transferring magnitude sequences with our Schmidt. The exposure times used were two minutes each. For instance, in a three-colour survey of S.A.8, magnitudes have been transferred from h and χ Persei about 10° away. On individual plates of 3 minutes exposure the scale over about eight magnitudes and the zero point differ by about 0·1 mag in each colour, down to the limit to which they were measured, i.e. 14^m. However, one would expect these differences to be randomly distributed, the systematic effects of flexure and declination being small in this case. The magnitude of one bright star in selected Area 8 was kindly measured by Mr. R. V. Willstrop using the pulse counting U.B.V.-photometer on the Cambridge 36-inch reflector. His result for U agreed with the mean of two plates within 0·04 mag; and for V it agreed with $\frac{1}{2}(R+G)$ from two R- and two G-plates within 0·03 mag.

On the other hand, transfers made with exposure times of the order of thirty minutes usually have scale and zero-point errors of several tenths of a magnitude. Some comment on this is made in the next section.

8. Conclusion

Preliminary experience of in-focus photometry with the Cambridge Schmidt telescope indicates that over the unvignetted field most of the error of a single magnitude measurement is due to photographic grain-fluctuation, and that for short exposures the field error is low, not exceeding a few hundredths of a magnitude in the unvignetted region, and that magnitude transfers can be accomplished to a few hundredths of a magnitude with certain precautions. In brief, the performance of this telescope in short-exposure photometry can be quite satisfactory.

This has not always been the experience with some other Schmidt telescopes, and it seems likely that the difference lies partly in the moderate dimensions and focal ratio of the Cambridge instrument (f/3·8; f = 163 cm). Under average conditions the diameter of the faintest ultraviolet images (about 25μ) is believed to be limited by seeing spread, because on rare occasions smaller diameters have been obtained. The G and R images, on the contrary, have never been obtained smaller than 25μ. These images are always less crisp than the ultraviolet ones, so that under fair seeing conditions it is thought that their diameter tends to be limited by emulsion spread.

It seems reasonable to suppose than an image limited by emulsion spread is less

sensitive to small disturbances than one limited by seeing. Perhaps this is supported by the figures in Table 3 of the present article; the measured U-magnitude is more sensitive to focus change than G or R. Of course, this comparison is certainly affected by wavelength-dependent optical and photographic effects, so that no final conclusion can be drawn from the present data; but a test could be made with different emulsions.

On long exposures the images tend to appear a little more diffuse, partly through reduced efficiency in guiding, and partly through photographic effects such as fogging by the sky brightness. Whether these render the magnitudes more susceptible to small disturbances of the image is at present a matter of speculation, but the disturbances are likely to be greater. The field errors for long exposures have not been studied yet, but they are likely to be different from those plotted in Figs. 9 and 10; and variable because of the movement of the plateholder already referred to. The variation of tilt of the plateholder with elevation has been shown to be negligible, but movement of the centre of the corrector plate relative to the centre of curvature of the mirror may cause an appreciable dependence of performance on zenith distance. Our experience with long-exposure transfers has been discouraging, and this agrees with the experience of other workers. Many contributors to the Draft Report of Commission 25 for the 1958 Congress of the International Astronomical Union have emphasized the need to set up magnitude scales *by photoelectric photometry in the actual areas being studied*, and not to attempt photographic transfers.

Now, although there appears to be no difficulty in doing photometry down to, say, the 15th magnitude with a Schmidt telescope of moderate size, making use of the wide field, and transferring magnitudes from standard sequences, many workers may wish to go several magnitudes fainter. There is little difficulty in obtaining the actual photographs: a large, fast Schmidt telescope will provide images certainly to the 20th magnitude, giving adequate plate-scale for photometry in many regions, and the wide field gives the astronomer sometimes several of the regions that interest him on one plate. (The advantage of a long focal-length in photometry, giving a large seeing-limited image, follows from equation (2). The larger number of grains reduces ΔD_{10}). When it comes to the photometry, however, the situation becomes difficult. First of all, it is generally found that a magnitude scale in one part of the plate cannot in general be transferred to another: differences in scale and especially in zero-point are found. These are usually attributed to tilt of the plateholder, as is explicable from Fig. 7. The magnitude changes plotted there refer to an f/3·8 telescope. For a faster system one would expect greater changes for a given error in focus. In principle it is possible to eliminate the effects of such a field error in the comparison of two regions in the field by reversing the telescope in such a way that the regions exchange position on the plate. However, this requires the field errors to remain invariant with respect to position on the plate on reversing. In practice they are almost certain to change, due to mechanical instability; and in any case the field error is not some absolute entity; it depends on the seeing and guiding.

As already mentioned, the conclusion has been reached by many observers that each region studied must have its own photoelectric magnitude sequences, so that in-focus photographic photometry becomes a matter of interpolation over only small distances on the plate. This, of course, simplifies the photographic measurement since each plate may have just one exposure, thus considerably lightening the task of identifying the stars in crowded fields. On the other hand, the setting up of

photoelectric sequences, even to the 18th magnitude, is a highly specialized technique which can be carried out at very few observatories at present. The problem then becomes one of securing the co-operation of an observatory equipped for this task, and this is almost certain to involve delays and frustration.

It seems then that there is a case for considering more critically the process of transferring magnitude sequences photographically. Evidently, the process fails because changes of image structure occur which are not accounted for in the measurement. The question of how to allow for them is one that has elicited much discussion with my associates here, especially Prof. REDMAN, Dr. LINFOOT and Dr. FELLGETT. Already a start has been made in the paper by H. F. WEAVER (1947) already referred to. It seems desirable to continue investigation along these lines rather than to write off photographic magnitude transfer with the Schmidt as a total loss.

In whatever way the problem of standard magnitude transfer is resolved, the moderate-sized or large Schmidt telescope provides a powerful instrument of mass photometry, so that by the use of modern irisphotometers and electronic computers the astronomer has a means of obtaining very extensive data, perhaps as rapidly as hand measurement can permit. This, then, is the *vista* before him.

REFERENCES

BECKER, W. and BIBER, C.	1956	*Z. Astrophys.* **41,** 52.
BECKER, W. and STEINLIN, U.	1956	*Z. Astrophys.*, **39,** 188.
COUDER, A.	1955	*Vistas in Astronomy*, ed. A. Beer, Vol. I, p. 372, Pergamon Press, London.
DEWHIRST, D. W. and YATES, G. G.	1954	*Observatory*, **74,** 71.
FELLGETT, P. B.	1958	M.N. **118,** 224.
HAFFNER, H.	1953	*Veröff. Sternwarte Göttingen*, Nr. 106.
	1956	*Mitt. Sternwarte Hamburg-Bergedorf*, **23,** Nr. 255.
JOHNSON, H. L. and MORGAN, W. W.	1953	*Ap. J.* **117,** 313.
LINFOOT, E. H.	1955	*Vistas in Astronomy*, ed. A. Beer, Vol. I, p. 351. Pergamon Press, London.
MEES, C. E. K.	1942	*The Theory of the Photographic Process*, 1st edn., p. 854. Macmillan, New York.
ROSS, F. E.	1936	*Ap. J.*, **84,** 241.
WEAVER, H. F.	1947	*Ap. J.*, **106,** 366.

The Excitation of Emission Lines in the Spectra of Early-Type Stars

B. E. J. Pagel

Royal Greenwich Observatory, Herstmonceux, Sussex

Summary

In a plane-parallel atmosphere, in radiative and local thermodynamic equilibrium, emission lines cannot normally arise. The presence of emission lines requires either a combination of thermal re-emission in the line with scattering in the continuum (Schuster-Underhill mechanism); or fluorescence; or scattering in a moving envelope. The various basic mechanisms for the production of emission lines are discussed: Of stars show the effects of selective fluorescence, while in WR, P Cygni and Be stars Zanstra's primary fluorescence mechanism appears to be operative. Rottenberg's effect of the scattering of stellar line radiation gives an elegant explanation of the P Cygni and Be profiles, but needs to be supplemented by a consideration of fluorescence effects before it can be applied to the higher members of a series.

The analysis of line intensities in WR and P Cygni stars according to Beals' Hypothesis is discussed in some detail, with special reference to the statistical equilibrium of hydrogen and helium in envelopes.

Finally the transfer of radiation in envelopes is discussed, with special reference to Sobolev's theory, in which self-absorption in a moving envelope is taken into account in a very elegant manner. It is concluded that Sobolev's theory offers much new insight, but is not universally applicable; for example, the measured ratio of Paschen to Balmer lines in the spectrum of P Cygni suggests that there is little self-absorption present in the higher members of the series.

1. Introduction

The presence of emission lines in the integrated spectrum of a star is a sign of the breakdown of one or more of the following assumptions that are commonly made in the theory of stellar atmospheres:

1. Radiative equilibrium
2. Hydrostatic equilibrium
3. Local thermodynamic equilibrium.

Of these assumptions, 3 may be expected to break down under a variety of conditions and 2 may be abandoned if there is sufficient evidence against it (e.g. if line profiles suggest turbulence, as in supergiants, or expansion, as in P Cygni and WR stars and novae). Assumption 1, on the other hand, is not to be abandoned lightly; while it is clearly inadequate to explain the high kinetic temperatures of the solar corona and the helium-emitting elements of the chromosphere, it has (when combined with Zanstra's fluorescence hypothesis) proved highly successful in accounting for the emission lines in gaseous nebulae and in the spectra of certain O, B and early A-type stars. However, the appearance of highly enhanced forbidden lines such as those of [Fe VII] in the spectra of novae (Thackeray, 1955) requires a coronal (i.e. non-radiative) excitation mechanism.

The emission lines, and certain narrow absorption lines that normally appear with them, are commonly assumed to arise in a shell or ring, analogous to a small

planetary nebula, surrounding the central star. The circumstantial evidence for this hypothesis is very strong in many cases, and consists of the following facts:

1. The weakness of absorption lines such as Mg II λ 4481, arising from subordinate levels with a strong transition to the ground state, together with the strengthening of lines arising from a metastable lower state, is consistent with a dilution factor of the order of 0·05 (STRUVE and ROACH, 1939). The presence of a dilution factor naturally assists the operation of the fluorescence mechanism.

2. The narrow "shell" absorption lines in Be spectra may be interpreted as showing a rotational broadening consistent with the dilution factor inferred from λ 4481 if it is assumed to arise in gas expelled from the (rapidly rotating) stellar surface with constant angular momentum, or perhaps with rotational velocity varying with distance according to KEPLER's Third Law.

3. The interpretation of P Cygni and WR profiles in terms of an expanding envelope is confirmed by the fact that similar profiles in the pre-maximum and principal spectra of novae are followed some weeks later by visible nebulosities with the same expansion velocities as shown by the principal spectra (MCLAUGHLIN, 1956). With the aid of suitable filters, similar nebulosities may be observed around certain P Cygni stars or slow novae such as AG Carinae (THACKERAY, 1956).

4. The rapid changes (sometimes occurring in a few hours or less) in the intensities and profiles of emission and "shell" absorption lines in Be spectra suggest violent ejection of material. Such material probably has a non-homogeneous structure and should be regarded as consisting of vast prominences; however, in the absence of detailed information as to its forms, it is convenient to treat the ejected material as though it had spherical (or, in the case of Be stars, circular) symmetry.

It is of some interest to analyse the distinction between the "shell" and the "star proper" from the viewpoint of the theory of stellar atmospheres. In the case of an expanding nova shell, or of the ring nebula in Lyra, there is no problem, since there is a region of negligible density between the star and the shell, so that the latter may be regarded as an isolated gas cloud illuminated by diluted stellar radiation. In the case of a star like P Cygni, however, material appears to have been continuously ejected at a more or less uniform rate since its nova outburst in 1600, so that the star must simply have an outer atmosphere whose density varies as the inverse square of the distance from the centre. The limb of the star is the layer whose transverse optical depth in the observable continuum is unity; below this layer we have a stellar atmosphere which can be legitimately discussed by the ordinary theory of radiative transfer, though it may well be necessary to allow for curvature by the methods of CHANDRASEKHAR (1934) and KOSIREV (1934), especially if the scale height is greatly increased by the expulsion process (breakdown of the hydrostatic hypothesis). KOSIREV showed that curvature could lead to a significant reduction in the colour temperature for a given spectral type, and thus accounted for the anomalously low colour temperatures observed in several early-type stars; this anomaly is nowadays usually attributed to interstellar reddening, but the "intrinsic reddening" suggested in the atmosphere of P Cygni and other stars (SHERMAN and MORGAN, 1939; BURBIDGE and BURBIDGE 1955) is probably due to KOSIREV's effect, perhaps in combination with an appreciable amount of continuous emission in an optically thin envelope.

Adapting an argument given by SOBOLEV (1947), whose work will be discussed in some detail later, it is possible to obtain a relation between the electron density

$N_e^{(0)}$ at the limb and the stellar radius R_*, assuming the density relation for a uniformly expanding shell,

$$N_e(R) = N_e^{(0)} \frac{R_*^2}{R^2}, \qquad \ldots\ldots (1)$$

and the predominance of THOMSON scattering by free electrons as the source of continuous opacity.[1] By definition of the limb, the horizontal optical depth at R_* is unity, i.e.

$$1 = \sigma_e N_e^{(0)} \int_{-\infty}^{\infty} \frac{R_*^2}{R_*^2 + y^2}\, dy = \pi \sigma_e N_e^{(0)} R_*, \qquad \ldots\ldots (2)$$

where σ_e is the cross-section per free electron and y is the distance from the limb to an element on a line of sight tangent to it. Hence

$$N_e^{(0)} = 6{\cdot}85 \times 10^{12} \left(\frac{R_*}{R_\odot} \right)^{-1}, \qquad \ldots\ldots (3)$$

where $R_\odot$ is the solar radius.

Above the limb, whether the atmosphere is expanding or not, the opacity is so small that continuous radiation longward of the Lyman limit undergoes very little further interaction with matter. Slightly shortward of the Lyman limit, the optical depth in the constant-ejection model is of the order of unity if the star is assumed to be a supergiant with ultra-violet brightness temperature 20,000°, though this value increases greatly for lower temperatures and luminosities. In any case, however, Lyman radiation becomes appreciably diluted before it has been absorbed and reemitted once or more; even when the Lyman optical depth is high, as in an H II region surrounded by an HI region (STRÖMGREN, 1939), the majority of the transfer processes take place in the narrow boundary zone between the two regions, since it is in this zone of relatively low ionization that the opacity is concentrated.

It follows that, in the greater part of the ionized region which is the source of line emission, the normal radiative transfer theory is inapplicable. The right approach is more likely to be along the lines of the Harvard Nebular Series (MENZEL, ALLER, BAKER, GOLDBERG, HEBB and SHORTLEY, 1937–1945). The detailed solution of the equations of statistical equilibrium has so far been carried out only for certain cases, and it is the purpose of this article to review the cases that have been discussed, together with their application to the quantitative observational material that is available, and to provide some suggestions for future work.

The physical origin of extended envelopes is discussed by STRUVE (1942). Probably material is driven out by the radiation pressure of Lyman α and Lyman continuum radiation (MIYAMOTO, 1953). The corresponding problem for planetary nebulae is discussed in some detail by ZANSTRA (1949); the physical situation in envelopes is analogous to that in the nebulae and will not be discussed here.

[1] For $N_e = 10^{12}$, in thermodynamic equilibrium, absorption at the heads of the Balmer and Paschen continua is less than 10 per cent of Thomson scattering if $T \geq 20{,}000°$.

2. Basic Mechanisms for the Production of Emission Lines

(i) *The Schuster-Underhill Mechanism*

This is the only mechanism that can lead to emission lines in a "normal" stellar atmosphere, i.e. in one that can be considered as being stratified in plane parallel layers and in which no fluorescence processes occur. The basic theory was given by Schuster (1905) in his classic paper on "Radiation through a Foggy Atmosphere". Imagine a uniform scattering medium of temperature T and optical depth τ illuminated from below by continuous radiation of intensity S in a certain wave-length range. The emergent intensity is given in the linear-flow approximation by

$$R = \frac{S}{1 + \frac{1}{2}\tau} \qquad \ldots\ldots (1)$$

and decreases steadily as τ is increased. Suppose now that line absorption takes place and that the excited state of the relevant atom is depopulated mainly by collisions, i.e. "pure absorption" occurs in accordance with Kirchhoff's law. If the optical depth is sufficiently large, the emergent intensity in the line is

$$R_\nu = \frac{2}{1 + \sqrt{\frac{1}{\epsilon}}} B_\nu(T)$$

and

$$R_\nu \to B_\nu(T) \text{ as } \epsilon \to 1, \qquad \ldots\ldots (2)$$

where $B_\nu(T)$ is the Planck function and ϵ is the ratio of absorption to the sum of absorption and scattering coefficients. In a stellar atmosphere, $B_\nu(T)$ is likely to be comparable with S, so that the combination of continuous scattering with selective absorption is capable of producing emission lines, even for quite small values of ϵ. Schuster considered the appearance of the spectrum for various values of τ and of the ratio of S to $B_\nu(T)$, and showed that, with atmospheres of suitable temperatures, it was possible to obtain a dark line, a bright line, or a bright reversal in the centre of a dark line. For temperature increasing with depth in a grey atmosphere, Schuster showed that the condition for emission is more easily fulfilled at longer wave-lengths and with lower temperature gradients, and concluded (wrongly) that the latter effect favoured emission in supergiant atmospheres.

Schuster's theory was rediscussed by Milne (1928) with the aid of the relationship between T and τ derived from radiative equilibrium theory. By this time it was generally agreed that the transfer of continuous radiation was by absorption, while that in line frequencies was by scattering, in most cases. Milne therefore concluded that Schuster's mechanism would be inoperative.

More recently, however, the realization of the importance of continuous Thomson scattering by free electrons in O and early B-type stars (especially in supergiants) has led to a rehabilitation of this mechanism (Underhill, 1949) for certain lines that may be considered to be formed by "pure absorption" if the gradient of $B_\nu(T)$ is not too steep. In a Milne-Eddington model with continuous absorption and

scattering coefficients κ_ν, σ and a selective absorption coefficient l_ν, Miss Underhill obtains the following condition for a line to appear in emission:

$$\frac{b_\nu^{(0)}}{b_\nu^{(1)}}\{R_2(\lambda_\nu)-R_2(\lambda_{\rm cont})\} > R_1(\lambda_{\rm cont})-n_\nu R_1(\lambda_\nu), \quad \ldots . (3)$$

where the Planck function is assumed given by the linear relation with monochromatic optical depth

$$B_\nu = b_\nu^{(0)}+b_\nu^{(1)}\tau_\nu,$$

$$\lambda_\nu = \frac{\kappa_\nu+l_\nu}{\kappa_\nu+l_\nu+\sigma}, \qquad n_\nu = \frac{\kappa_\nu+\sigma}{\kappa_\nu+l_\nu+\sigma}, \quad \ldots . (5)$$

and the functions R_1, R_2, which involve λ_ν and the first and second moments of Chandrasekhar's function $H(\mu)$ (Chandrasekhar, 1947), are tabulated by Miss Underhill. If near the stellar surface we assume the relationship between temperature and mean optical depth

$$T^4 = T_0^4(1+\tfrac{3}{2}\bar{\tau}), \quad \ldots . (6)$$

we have

$$\frac{b_\nu^{(0)}}{b_\nu^{(1)}} = \frac{8}{3}\frac{kT_0}{h\nu}(1-e^{-h\nu/kT_0})\frac{\kappa_\nu+\sigma}{\bar{\kappa}+\sigma}, \quad \ldots . (7)$$

and we can then apply condition (3) to a model atmosphere, if a suitable one is available.

It is of some interest to apply (3) to the model calculated by Miss Underhill herself for an O 9·5 V star with surface temperature $T_0 = 25{,}200°$ (Underhill, 1950). In this model, $(\kappa_\nu+\sigma)/(\bar{\kappa}+\sigma) \simeq 1{\cdot}0$ at $\lambda = 1$ micron, so that approximately

$$\frac{b_\nu^{(0)}}{b_\nu^{(1)}} \simeq 4{\cdot}8\,\lambda, \quad \ldots . (8)$$

where λ is expressed in microns. For a strong absorption line, $\lambda_\nu \simeq 1$, $n_\nu \simeq 0$, while $\lambda_{\rm cont}$ is about 0·7. Hence (3) becomes

$$4{\cdot}8\lambda\{R_2(1)-R_2(0{\cdot}7)\} > R_1(0{\cdot}7),$$

which, on substituting the R's, gives

$$\lambda > 1{\cdot}8 \text{ micron}. \quad \ldots . (9)$$

In this model, therefore, no emission is to be expected except in the far infra-red. In a supergiant, on the other hand, the electron pressure may be down by a factor of 10 and κ_ν/σ is likely to go down by a similar factor, so that $\lambda_{\rm cont}$ will be about 0·2. In this case, our condition becomes

$$\lambda > 4500\,\text{Å} \quad \ldots . (10)$$

so that one might expect emission in Hα and Hβ in O9 supergiants, extending to higher members of the series for earlier spectral classes. However, the operation of

the mechanism requires a strong line formed mainly in accordance with the "pure absorption" mechanism, and detailed model calculations, combined with a careful study of the emission profiles in, say, Of stars are needed to verify its importance. In fact, Of emission lines usually show selectivities suggesting the action of fluorescence (see (v)).

(ii) *Cyclic Processes in a Plane-parallel Atmosphere*

The non-zero central intensities of Fraunhofer lines are inconsistent with the hypothesis of pure coherent scattering. Their explanation demands one of the following: de-excitation by collisions, which is normally unimportant for an upper state with low intrinsic lifetime; redistribution in frequency within the line; or interlocking with other lines and with the continuum (Woolley, 1934; Strömgren, 1935). This latter effect is based on the theorem of Rosseland (1926) that, in the presence of dilute black-body radiation, cycles occur more frequently in the sense of absorption of high-frequency radiation followed by emission of low frequencies than in the reverse sense; such cycles, especially those involving ionisation from the ground state followed by recombination to excited states, form the primary fluorescence mechanism which is almost certainly the main source of emission lines in early-type stellar atmospheres. The Rosseland cycle represents the tendency of the diluted radiation to interact with matter in such a way as to approach more closely to thermodynamic equilibrium by reducing its own colour temperature; its effect is naturally most marked in extended shells and planetary nebulae where the dilution factor W is very small, but it can also operate in the surface layers of a normal star such as the Sun, which can be regarded as being illuminated by colour-temperature radiation diluted by a factor of 2 or 3 (the exact value depending on the spectral gradient in the relevant ionizing continuum, which is of course poorly known).

Woolley studied the special case of Hα and Hβ in the solar spectrum, taking into account the mutual influence of 5 levels (the fifth being in the continuum) of a schematized hydrogen atom. Woolley showed that the minimum residual intensity in a line is $(\alpha - 1)$, where α is the ratio of downward to upward transitions that would have occurred in the presence of the diluted radiation undiminished by an absorption line. The excess of α over unity is primarily due to the excess of recombinations to higher levels (a fixed quantity if ionization equilibrium at a given density and kinetic temperature is assumed) in comparison with absorbing transitions (the number of which is proportional to W), and is inversely proportional to W. This "minimum" residual intensity differs from the observed value by an amount depending on the local intensity of Lyman-α radiation, this dependence being strong for Hα and much weaker for Hβ. When this effect was taken into account, Woolley obtained good agreement with the observed central intensities of Hα and Hβ on the solar disc, and with the sensitivity of Hα to local conditions as displayed in spectroheliograms, but the results require re-examination in the light of more recent values of the continuous absorption coefficient and of the fact that Lyman α itself appears as an emission line; the latter phenomenon has been discussed recently as a fluorescence effect (de Jager, 1956), but this discussion depends on calculations of the population of each level (Giovanelli, 1948) based on the physically unrealistic assumption of negligible line absorption in all series in the chromosphere.

Woolley's discussion led to the conclusion that, in a large tenuous envelope with negligible continuous absorption and a sufficiently small value of W, one could meet

the necessary condition for emission, $\alpha \geqslant 2$, and that high stellar temperature would assist the process.

A generalized discussion of the effect of electron captures on the residual intensities of resonance lines was given by B. STRÖMGREN (1935) for a schematized atom with a ground state, a single excited state, and a continuum. By considering the statistical equilibrium of the excited level, STRÖMGREN showed that recombinations led to an extra term in the transfer equation, formally equivalent to a collision term, thus increasing the effective value of the parameter ϵ. Strömgren obtained for the population N_k of the excited level

$$\frac{N_k}{N_k^{(0)}} = \xi_k = (1-\epsilon)\frac{J_\nu}{B_\nu(T)} + \epsilon, \qquad \ldots . (11)$$

where

$$\epsilon = \frac{\sigma_{k1}N_e + C_{kf}}{A_{k1} + \sigma_{k1}N_e + C_{kf}}, \qquad \ldots . (12)$$

the superscript $^{(0)}$ refers to local thermodynamic equilibrium, J_ν is the intensity of radiation at the line frequency averaged over direction, σ_{k1} is the collisional deactivation coefficient, N_e the electron density, C_{kf} the rate of photo-ionization from the excited state, and A_{k1} the spontaneous transition rate to the ground state. The corresponding source function in the transfer equation, S_ν, is given by

$$(1+\eta)S_\nu = \eta(1-\epsilon)J_\nu + (1+\epsilon\eta Q)B_\nu(T), \qquad \ldots . (13)$$

where η is the ratio of selective to continuous extinction coefficients and the correction factor Q, which always exceeds unity in a dilute radiation field, is given by

$$Q = \left(\frac{C_{1f}}{C_{kf}}\right) \Big/ \left(\frac{C_{1f}}{C_{kf}}\right)^{(0)}. \qquad \ldots . (14)$$

G. R. and E. M. BURBIDGE (1952) used STRÖMGREN's method to examine the central residual intensities of the first three Balmer lines, extending STRÖMGREN's equations of statistical equilibrium to include transitions to and from level 1 and to take some account of stimulated emissions. The transfer equation included the effect of continuous electron scattering and was solved by CHANDRASEKHAR's method. For this purpose, Q was replaced by

$$Q' = \frac{\xi_1}{\xi_2}Q, \qquad \ldots . (15)$$

but, to avoid the necessity of solving two linked transfer equations simultaneously, ξ_1/ξ_2 was assumed to be unity (corresponding to L.T.E. in the relative populations of the Lyman and Balmer levels). BURBIDGE and BURBIDGE obtained the following condition that the intensity in the line centre exceed that in the neighbouring continuum:

$$\frac{b_\nu^{(0)}}{b_\nu^{(1)}}(Q'x - y) > z - Q'w, \qquad \ldots . (16)$$

where x, y, z, w are constants (in this approximation) for a given atmosphere at a

given wave-length involving the ratio κ_ν/σ of continuous absorption to scattering, ϵ, γ_l and the first and second moments of $H(\mu)$.

In agreement with Miss UNDERHILL, the BURBIDGES concluded that no emission can occur in a line formed by pure scattering, since Q' only occurs with the factor ϵ in (13)—this hypothesis gives the fluorescence processes no chance to influence the result. When $\epsilon > 0$, however, the conditions for emission are somewhat easier to fulfil than in Miss UNDERHILL's case. For stellar temperatures between 20,000° and 30,000° and with plausible values of ϵ, they showed that $Q'\epsilon$ must exceed about 4 for emission to occur in Hα, Hβ or Hγ in dwarfs or giants, the limit diminishing slowly with increasing temperature; such values might be expected from (15) if, owing to departures from equilibrium, $\xi_1 > \xi_2$, but not otherwise.

These results suggest that the SCHUSTER-UNDERHILL mechanism is not very likely to work in most cases (since ϵ is too small), and that the fluorescence mechanism normally requires an extended shell (with dilution factors appreciably less than $\frac{1}{2}$) to produce emission lines except at very high temperatures ($> 40{,}000°$).

(iii) *Fluorescence in an Extended Shell*

The "primary fluorescence mechanism" was first put forward in a definite and quantitative form by ZANSTRA (1927, 1931a) for nebulae; stellar radiation on the shortward side of the Lyman limit causes ionization of hydrogen in the nebula; subsequent recombinations to excited levels give rise to emission lines in all the subordinate series and continua, notably the Balmer series. Owing to dilution in the nebula or shell, the neutral hydrogen is almost entirely in the ground state; hence, in accordance with ROSSELAND's theorem, high-frequency radiation is degraded into low frequencies without being replaced by the reverse cycle (ionization from excited states followed by recombination to the ground state). A similar mechanism accounts for many of the other permitted emissions that are observed, e.g. those of helium and of ionized carbon, nitrogen, oxygen, silicon and iron, provided that the central star is hot enough; the emission of each ion comes from a region in which the atom concerned is predominantly in the next higher state of ionization.

Each ultra-violet quantum that is absorbed causes the emission of one quantum in a line or in the continuum of the first subordinate series (such as the Balmer series), provided that the lines of the fundamental series are heavily absorbed, while those of the subordinate series escape with negligible absorption. In this case, a Lyman line is ultimately degraded into Lyman-α and a Balmer line (in some cases together with members of higher series). Making the additional assumption that virtually all quanta in the Lyman continuum are absorbed, ZANSTRA was able to apply this process to the computation of the temperature of the central star from the equivalent widths of the nebular Balmer lines in units of the underlying stellar continuum.

Elementary considerations based on this process lead to two important results: a relation between the stellar temperature T_* and the electron temperature T_e in the nebula (assumed for the moment to consist of pure hydrogen) and a relation between T_*, the density of the nebula, and the maximum size of the ionized region. An expression for the first relation was obtained by EDDINGTON (1926), who calculated the mean kinetic energy of electrons expelled from the ground state of hydrogen by stellar radiation in the Lyman continuum. The collisions of the expelled electrons

with one another and with the atoms and ions rapidly set up a Maxwellian velocity distribution with kinetic temperature

$$T_e \simeq \tfrac{2}{3}T_*(1-kT_*/\chi_1), \qquad \ldots . (17)$$

where k is Boltzmann's constant and χ_1 the ionization energy from the ground level. More elaborate calculations of the electron temperature in an optically thin hydrogen nebula were carried out later by BAKER, MENZEL and ALLER (1938), who took the various absorption and emission processes into account in detail and considered the energy balance; their results gave T_e of the same order of magnitude, but somewhat higher. The results for a hydrogen nebula are in fact inapplicable to planetary and diffuse nebulae, because these are always cooled to temperatures close to 10,000° by the emission of collisionally excited lines, mainly of nebulium (ALLER, 1956a). However, the shells surrounding emission-line stars are normally of such a high density that no forbidden lines appear; in the absence of permitted transitions from low-lying levels, collisional excitations are negligible (see below, Section 3 (vi)) and the results for a pure hydrogen nebula may therefore be expected to apply.

The extent of the ionized region is limited by the absorption of ionizing radiation by the residual amount of the next lower ion (e.g. neutral H in the H II region in which the H I spectrum is emitted by the recombination of electrons and protons). A simple method of estimating the maximum extent of an H II region was first given by ZANSTRA (1931b), using the condition that the optical depth is of the order of unity, i.e.

$$\int_{R_*}^{R_1} \alpha N_1 dR = 1, \qquad \ldots . (18)$$

where R is distance from the centre of the star (radius R_*), α the absorption cross-section per neutral hydrogen atom shortward of the Lyman limit (which may be placed equal to its value at $\lambda 912$, $6{\cdot}3 \times 10^{-18}$ cm^2), and N_1 the density of neutral hydrogen atoms, given by the ionization equation (neglecting absorption)

$$N_1 = N_e^2 \frac{4R^2}{{R_*}^2}\left(\frac{h^2}{2\pi mkT_*}\right)^{\frac{3}{2}}\left(\frac{T_*}{T_e}\right)^{\frac{1}{2}} e^{\chi_1/kT_*} \qquad \ldots . (19)$$

(PANNEKOEK, 1926), where N_e is the electron density (nearly equal to the total heavy-particle density N in the H II region). From (18), (19), neglecting R_* in comparison with R, one obtains

$$\frac{R_1}{R_*} = \left[\frac{1}{N^2}\left(\frac{R_*}{R_\odot}\right)^{-1} 10^{21{\cdot}62-68538/T_*}\left(\frac{T_e}{T_*}\right)^{\frac{1}{2}}\right]^{\frac{1}{3}}, \qquad \ldots . (20)$$

where $R_\odot$ is the solar radius. More elaborate treatments by ZANSTRA (1932) and STRÖMGREN (1939), in which the absorption factor $e^{\int \alpha N_1 dR}$ was taken into account in the ionization equation (19), led to exactly the same expression (20) for the extent of the ionized region. The identity of the two results is due to the fact that a co-operative phenomenon occurs at the point where the optical depth obtained from the simple expression (18) reaches unity; here the amount of neutral hydrogen becomes so large that the Lyman radiation is absorbed very rapidly and ionization virtually ceases in a very short distance.

This concept of stratification into successive limited shells whose stages of ionization decrease with increasing distance from the star is in good agreement with the observed stratification in the more symmetrical planetary nebulae (BOWEN, 1928; ALLER, 1956b). Analogous observations in the atmospheres of Wolf-Rayet stars will be discussed below (Section 3 (i)).

(iv) *Line Scattering in an Extended Moving Shell*

Coherent scattering in a stationary spherical atmosphere has been discussed by Miss UNDERHILL (1948), who showed that curvature of the atmosphere tends to reduce the central absorption. The simplest case is that of a thin spherical shell surrounding a star of much smaller size, where it is clear that scattering produces no absorption

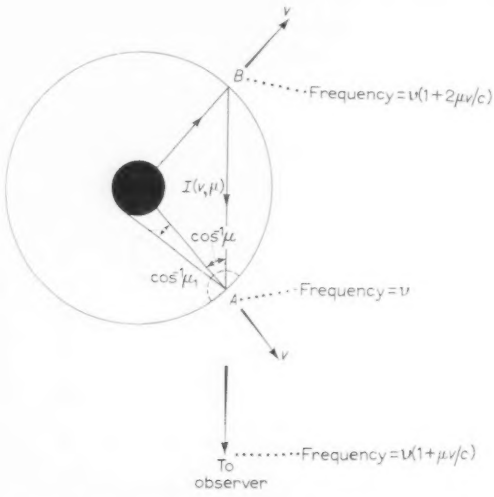


FIG. 1. Scattering of line radiation in a moving shell (according to ROTTENBERG).

line at all unless an appreciable portion of the shell is occulted by the star. If the shell is moving, the total equivalent width of the line remains zero, but its energy is redistributed among absorption and emission components, as has been shown for the particular case of an expanding envelope by ROTTENBERG (1952).

ROTTENBERG considers the idealized case of a geometrically thin spherical shell with radius R_1 and optical thickness $\tau(\nu)$ expanding with velocity v around a star of radius R_* that emits a continuous spectrum of specific intensity F (see Fig. 1). If $I(\nu, \mu)$ be the intensity of stellar radiation striking the inner surface of the shell at an angle $\cos^{-1}\mu$ to the normal, ν being the frequency seen by an observer moving with that portion of the shell, the net flux scattered into all directions by unit area is

$$\int_{\substack{\text{inward}\\ \text{half-sphere}}} I(\nu, \mu)\{1 - e^{-\tau(\nu)/\mu}\}\mu d\omega = 2\pi \int_0^1 I(\nu, \mu)\{1 - e^{-\tau(\nu)/\mu}\}\mu d\mu. \qquad \ldots . (21)$$

This flux can also be represented by the expression

$$\int_{\text{sphere}} S(\nu)\{1-e^{-\tau(\nu)/|\mu|}\}|\mu|d\omega = 2\pi S(\nu)[1-2E_3\{\tau(\nu)\}], \quad \ldots\ldots (22)$$

where $S(\nu)$ is the source function representing scattered radiation and the exponential integral function $E_n(x)$ is defined by

$$E_n(x) = \int_1^\infty e^{-xy}y^{-n}dy. \quad \ldots\ldots (23)$$

Assuming coherence of scattering, $S(\nu)$ can now be found as a function of I by equating (22) to (21). The introduction of non-coherent scattering will not alter this result if it is assumed that the line profile for emission is identical with that for absorption.

Now radiation striking one point of the shell A at an angle $\cos^{-1}\mu$ and with frequency ν is seen by an observer at rest relative to the star with frequency

$$\nu' = \nu(1+\mu v/c) \quad \ldots\ldots (24)$$

and by an observer moving with another portion of the shell B (which emitted it before it was scattered at A) with frequency

$$\nu'' = \nu(1+2\mu v/c). \quad \ldots\ldots (25)$$

For $\mu_1 < \mu \leqslant 1$, where $\mu_1 \equiv (1-R_*{}^2/R_1{}^2)^{\frac{1}{2}}$ is the cosine of the semi-angle subtended by the star at a point on the shell,

$$I(\nu, \mu) = F; \quad \ldots\ldots (26)$$

for larger angles of incidence,

$$I(\nu, \mu) = S(\nu'')\{1-e^{-\tau(\nu'')/\mu}\}. \quad \ldots\ldots (27)$$

Hence the emissivity is finally given by the integral equation

$$\begin{aligned} S(\nu)[1-2E_3\{\tau(\nu)\}] = &\int_{\mu_1}^1 F\{1-e^{-\tau(\nu)/\mu}\}\mu d\mu + \\ &+\int_0^{\mu_1} S(\nu'')\{1-e^{-\tau(\nu'')/\mu}\}\{1-e^{-\tau(\nu)/\mu}\}\mu d\mu. \end{aligned} \quad \ldots\ldots (28)$$

Numerical solution of (28) leads to the calculation of the line contour by summing three components: the star as seen through the envelope, the backward half of the envelope as seen through the forward half, and the forward half, or "disc". In a frame of reference that is stationary relative to the star, the emergent flux is

$$\mathcal{F}_v = F\int_{\mu_1}^1 e^{-(1/\mu)\tau\{\nu(1-\mu v/c)\}}\mu d\mu + \int_0^{\mu_1} S\{\nu(1+\mu v/c)\}\{1-e^{-(1/\mu)\tau\{\nu(1+\mu v/c)\}}\}e^{-(1/\mu)\tau\{\nu(1-\mu v/c)\}}\mu d\mu +$$

$$+\int_0^{\mu_1} S\{\nu(1-\mu v/c)\}\{1-e^{-(1/\mu)\tau[\nu(1-\mu v/c)]}\}\mu d\mu = \mathcal{F}_1+\mathcal{F}_2+\mathcal{F}_3. \quad \ldots\ldots (29)$$

The resulting profile depends on μ_1 and on β, the ratio of expansion velocity to the Doppler broadening parameter, which latter includes the effects of non-uniform expansion velocity as well as the usual thermal and turbulent components. For $\beta \geqslant 3{\cdot}3$, pure scattering gives a broad undisplaced emission component with a central reversal, in addition to a violet-displaced absorption line. $\beta = 1$ gives a simple emission component, while $\beta = 10$ gives a doubled emission component so wide that it is hardly noticeable. The physical reason for this is that the forward and rear halves of the shell separate into two sources that are optically thick for their own radiation but not to each other's owing to the Doppler shift (leakage effect),[2] if β is large. Each half-shell radiates like a surface in which a given interval μ to $\mu + d\mu$ in the cosine of the angular distance from the centre of the disk corresponds to an interval of frequency $\nu_0 \pm \mu v/c$ to $\nu_0 \pm (\mu + d\mu)v/c$; as usual, the contribution of each element to the net flux is weighted by the factor μ, so that the maxima occur near the extreme frequencies $\nu_0 \pm v/c$ and one obtains an intensity minimum at the undisplaced line frequency. The maximum on the violet side (coming from the forward half of the shell) is actually the lower one because of the lowering of the continuous background by the violet-displaced absorption of the shell. The weighting of radiation from each portion of the star by the factor μ leads to an additional violet displacement in the minimum of the absorption-line profile, which is therefore unsymmetrical (cf. Section 4(ii)).

Rottenberg proceeds to discuss the more complex case in which line radiation is generated by the primary fluorescence mechanism in an optically thin shell of radius $R_2 (\leq R_1)$ and which leads to an additional term in the expression for the flux (28). Here the simplest cases are those in which $R_* = R_2 \ll R_1$, where the previous theory must simply be applied to a continuous stellar background modulated by an emission line, and that in which $R_* \ll R_2 = R_1$. Both cases lead to the central reversal in the emission component; the absorption component is lifted bodily above the continuous background in the first case, and filled in by emission in the second. Intermediate cases correspond very closely with the various types of P Cygni profile given by Beals (1950); the great merit of Rottenberg's elegant treatment is that the introduction of the scattering effect removes the necessity for special velocity distributions to account for the observed profiles.

A further point of great interest pointed out by Rottenberg is the correspondence between the observed sequence of profiles of $H\alpha - \gamma$ in such stars as HD 190603 (Beals, 1950) and the calculated sequence for increasing occultation by the central star. This affords direct evidence of *stratification* in the shell, the higher lines of the Balmer series being mainly emitted in regions closer to the star than the lower ones. Such stratification should be taken into account in discussions of the Balmer decrement even in cases where scattering can be neglected; however, fluorescence among the higher members of the series is likely to lead to the same effect.

It would be of the greatest interest to carry out a similar analysis for a rotating envelope with a view to calculating the profiles of Be spectra, certain features of which—especially the "red-violet asymmetry" of McLaughlin (1931, 1933)—are already produced in Rottenberg's case $R_2 = R_1$ without taking rotation into account.

The main limitation of Rottenberg's theory is its reliance on pure scattering;

[2] This term is taken from the stimulating report by Kourganoff (1958).

when the optical depth of the scattering shell is large, fluorescence will tend to increase the intensity of Hα and reduce that of higher members of the series, provided that the optical depths in the Paschen and higher series are low. The steep Balmer decrements observed in many cases (such as HD 190603) suggest that this process is going on; it is unlikely that stratification can entirely account for the disappearance of the *scattered* emission components of the higher Balmer lines (since the scattering layers must be substantially the same for all of them), although it will naturally help to steepen the decrement for the components emitted by the primary fluorescence mechanism.

(v) *Selective Fluorescence*

Bowen (1934, 1935) pointed out that seven of the ten O III lines appearing in planetary nebulae originate either from the $3d\,^3P_2$ state, or from states directly attainable by a downward transition from it; while other lines of the same and nearby multiplets (some lines of which are actually stronger in the laboratory than those observed in the nebulae) are missing. This fact, together with the presence of [O III] lines, virtually rules out recombination as a source of the O III spectrum and Bowen pointed out the coincidence between the resonance transition $(2p)^2\,^3P_2 - 2p3d\,^3P_2$ of O III, λ 303·80, and that of He II ("Lyman α") at λ 303·78. Owing to the finite width and displacement of the line profile, resonance quanta emitted by recombination of He II may be absorbed by O III ions (leading to fluorescence) as well as scattered by He II itself; ultimately each He II quantum is converted into one or more O III quanta, after which only [O III] appears. The occurrence of Bowen's selective fluorescence mechanism is confirmed by the very close correlation between the appearance of O III lines and λ 4686 of He II. The same mechanism also accounts for the N III lines λλ 4097, 4634, 4641 through the coincidence between the transition $2p\,^2P_2 - 3d\,^2D_{2,3}$ (λ 374·44) of N III and the resonance line of O III λ 374·436 emitted by cascade after excitation by the He II line; furthermore, He II λ 4686 may itself be emitted when He II in the metastable $2s$ state (to which it is excited by the high density of λ 304 radiation) is raised to the $4p$ state by absorption of Lyman α, provided that there is a relative velocity of 120 km/sec between the H and He II atoms.

Similar effects presumably cause the selectivities observed in certain stars with emission lines. In P Cygni, for example, the Si III triplet $4p\,^3P^o - 4d\,^3D$ near λ 3800 appears purely in emission, while the triplet $4s\,^3S - 4p\,^3P^o$ near λ 4560 appears mainly in absorption with just one emission component on the red side of λ 4552 (Struve and Roach, 1939). Similar behaviour is shown by the ultra-violet members of the same multiplet and also by the singlet transitions involving $4p\,^3P^o$ (Swings and Struve, 1940). The population of the $4p\,^3P^o$ and $4p\,^1P^o$ multiplets, common to both transitions, must be abnormally low, no doubt because the P^o states cannot combine directly with the ground level. Similarly, Of stars show a selectivity in the emissions of N III (which here behaves very similarly to Si III in P Cygni), N IV, N V, C III, Si IV, He I, He II that is apparently uncorrelated with either excitation potential or metastability (metastable levels are normally enhanced in population in the presence of diluted radiation); this selectivity becomes less marked as the intensity of emission increases and seems to have disappeared altogether in WR envelopes (Swings, 1942). An exception to this is V444 Cygni (see Section 3(ii)).

These selectivities are discussed very thoroughly by SWINGS and STRUVE (1942), whose conclusions we shall now summarize. The relative importance of recombination and selective fluorescence in exciting radiation of an ion A is proportional to

$$N_{A^+}N_e/(N_A I_*)$$

where I_* is the intensity of stellar exciting radiation: and N_A, N_{A^+}, N_e are the number densities of A, A^+ and electrons, respectively. If A is more abundant than A^+, selective fluorescence will tend to predominate provided that suitable emission lines (or a strong enough continuum, since the geometrical dilution factor is high, of the order of 0·1) are present. In the case of He I, fluorescence favours the singlets, since the ground state is $1s\,^1S$ and intercombinations are absent, provided that radiation is present in the exciting range $\lambda\lambda$ 500–580. Thus singlets (e.g. λ 4388) are weak in the high-excitation objects where He^+ predominates and becomes about equally strong with the triplets (e.g. λ 4026) in cooler objects such as P Cygni. Pure geometrical dilution, on the other hand, favours the triplets (STRUVE and WURM, 1938). Essentially, the high optical thickness in resonance lines of the singlet series produces a high radiation density in these lines compared with that of the diluted continuum, so that they are able to induce Bowen fluorescence processes in the outer parts of the shell. Similarly, λ 4686 of He II is probably produced by the same mechanism as in nebulae.

On the other hand, N III (λ 4634 and λ 4641) and C III (λ 5696) display selectivities in the later Of stars that can only be accounted for by a modified form of BOWEN's mechanism. In the earlier Of stars, the weakness of the N III absorption lines suggests that N III emission is due to recombination, and the selectivity is much less marked. In the late Of stars, SWINGS and STRUVE suggest that N^{++} is excited to the $3d\,^2D$ level by absorbing continuous stellar radiation near λ 374, from which $\lambda\lambda$ 4634, 4640 ($3p\,^2P^o - 3d\,^2D$) will be emitted in accordance with ROSSELAND's theorem, as in the case of He I. Since the odd level $3p\,^2P^o$ cannot be excited directly from the ground state, it is populated only by these emissions and its population is low because it has numerous downward transitions; hence the lines $\lambda\lambda$ 4097, 4103 ($3s\,^2S - 3p\,^2P^o$) appear in absorption, in contrast to the emission by successive cascade that one might expect. Exactly the same applies to Si III in P Cygni and, *mutatis mutandis*, to many other cases. For the C III emission λ 5696, however, Miss UNDERHILL (1958a) suggests that helium radiation, λ 304, ejects a $2s$ electron from C^+ in the $1s^2 2s^2 3d$ state, leaving it in the appropriate upper level $3\,^1D_2$.

A further departure from the situation in planetary nebulae is indicated by the absence, in many Of stars, of O III emission; perhaps BOWEN's mechanism is here *prevented* from operating by a Doppler shift due to the expansion of the shell, as a result of which most excitation must occur through the absorption by each atom of its own exciting frequency (suitably Doppler shifted) from the stellar continuum. This process is naturally much more efficient here than in nebulae because the geometrical dilution factor is high. One might still expect O III to be excited, in the same way as N III, by absorbing $\lambda\lambda$ 300–345 from the continuum, but for some reason this does not appear to occur, maybe because of the existence of strong intercombinations. Where O III emission does occur, it is attributed by SWINGS and STRUVE to recombination. However, Miss UNDERHILL (1958b) has shown that the O III emission resulting from BOWEN's mechanism would be spread over so many transitions that

it would be very inconspicuous; she therefore proposes the explanation of the N III emission by a direct application of Bowen's cycle.

Oke (1954) has argued that the emission lines of Of stars arise in the "stellar atmosphere" rather than in a "shell",

(a) because the N III lines appear to be self-absorbed, requiring $W \to 1$;

(b) because the half-widths of the absorption and emission profiles are comparable even when the rotational velocity is high. On the other hand, the emission may be considerably red-shifted with respect to the absorption, a feature suggesting an expanding envelope in which, from (a), a significant degree of scattering occurs (Rottenberg's case $R_* \ll R_2 = R_1$).

This being so, Oke's argument loses is force, since it only indicates that the emission and absorption lines are formed in the same region. Owing to self-absorption, the geometrical dilution factor is not applicable to line radiation. Oke argues further that the stellar radiation does not contain enough high-energy quanta to produce the emission; this argument is very sensitive to the temperature assumed, since 0·1 Å of the continuum (or an emission line of this equavalent width, corresponding to only ± 35 km/sec) at λ 374 of a black body at 40,000° contains enough quanta to excite 0·5 equivalent angstrom of N III emission at $\lambda\lambda$ 4634, 4641 (about the order of magnitude observed by Oke before correcting for underlying absorption lines—a correction which is self-cancelling provided that the shell is large enough for occultation by the central star to be negligible). Although nitrogen in the radiation field of a black body at this temperature should be predominently in the N^{3+} and N^{4+} states, it is reasonable to assume that the ionizing continuum for $N^{2+}(\lambda \leqslant 264)$ is greatly depleted through photo-ionization of N^{2+}, C^{2+} ($\lambda \leqslant 259$) and He^{+} ($\lambda \leqslant 228$) before it reaches the visible layers of the atmosphere (Swings and Struve, 1942). Hence we prefer the explanation in terms of fluorescence to Oke's suggestion of mechanical excitation in the presence of a cooler star ($T_* = 30{,}000°$), which completely fails to account for the selectivity. A complete theory of line formation in these cases should discuss a combination of fluorescence with Rottenberg's line scattering mechanism; such a discussion offers a most attractive opportunity to account for the existence of a series of related absorption lines, some of which are accompanied by emission while others are not.

3. The Analysis of Emission Spectra

(i) *Beals' Hypothesis: Recombination in a Transparent Shell*

A simple and fruitful assumption in the analysis of novae, Wolf-Rayet and P Cygni stars has been Beals' hypothesis (Beals, 1934, 1940) that the mechanism of emission is identical with that suggested by Zanstra for gaseous nebulae and that the emissions of successive stages of ionization in the WR, nova and P-Cygni type spectra are stratified in the manner discussed for nebulae by Bowen (1928). Owing to the progressive absorption of ionizing radiation by successive layers of the gas, the highest frequencies are found nearest to the central star and successively lower degrees of ionization are encountered as one goes further out into the envelope. While stratification in the stellar envelopes is not directly observable, as it is in planetary nebulae, there is strong circumstantial evidence for it in the existence of emission lines from species of widely differing ionization potential (e.g. N V, I.P. 97v,

and He I, I.P. 24v, in WR stars; and Si II, I.P. 6v, and He II, I.P. 54v, in P Cygni stars) and in the increase in line width in a single spectrum as the ionization potential diminishes. In P Cygni stars, further evidence for stratification is provided by the decrease in the emission : absorption intensity ratio with increasing ionization potential. The flat-topped line profiles in many WR and nova spectra are consistent with uniform radial expansion in a transparent spherical envelope, which could be either a full sphere or a thin shell; the increase of line width with diminishing ionization potential then requires an outward acceleration of the moving material. Discussions of the line profile in the presence of velocity and emission gradients by GERASIMOVIČ (1933), CHANDRASEKHAR (1934b) and WILSON (1934) turn out to be difficult to apply in practice because they involve too many unknown parameters; however, it may prove possible to obtain details of the degree of outward acceleration by comparing the line widths of elements in successive degrees of ionization. For this purpose, it should be possible to assume that each ion is concentrated in a shell in which the velocity does not vary significantly. On the other hand, line profiles without a flat top, as in P Cygni and some WR stars, require either a velocity gradient in the shell occupied by one ion, or a high degree of random motion, or self-absorption.

BEALS' hypothesis has an immediate application to the determination of the temperatures of emission-line stars by ZANSTRA's method (BEALS, 1940; BEALS and HATCHER, 1948; BEALS and ROTTENBERG, 1956). This is particularly valuable because, at the high temperatures of these stars, even a small amount of interstellar absorption or of continuous emission in the envelope introduces a large error in colour temperatures derived from measurements in the photographic region. On the other hand, the temperatures derived by ZANSTRA's method are subject to some uncertainty for the following reasons:

(a) Stars do not radiate like black bodies in the far ultra-violet.
(b) The Lyman and corresponding continua may not be completely absorbed.
(c) The emission lines measured in the observable region do not represent all the quanta emitted by fluorescence.
(d) The emission lines may be cut down by self-absorption.
(e) Selective scattering and fluorescence may occur.

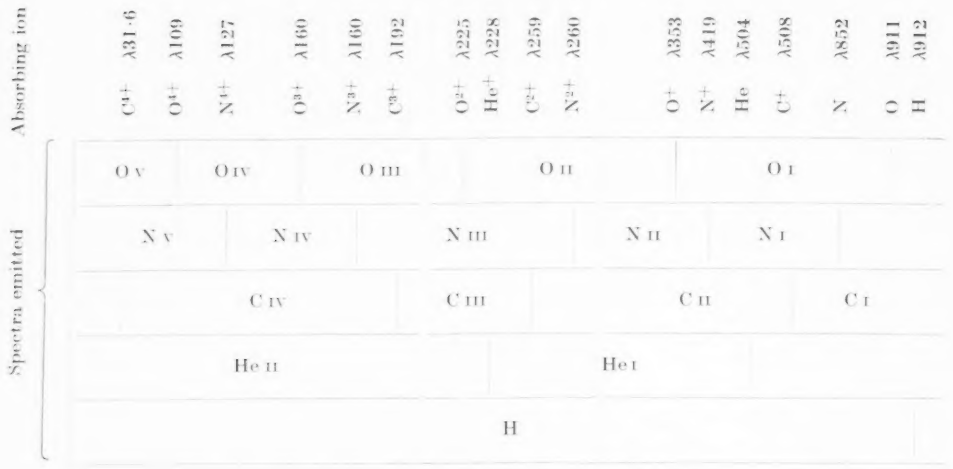
Of these difficulties, (a) is probably the most serious. In principle, the difficulty might be solved by an extended programme of model-atmosphere computations such as has been begun by Miss UNDERHILL (1958b); such a programme, however, would not only be very lengthy but also be open to the objection that fluorescence processes would occur to some extent in the stellar atmosphere itself, the separation of which from the envelope is inevitably somewhat artificial. A list of relevant emission lines in the far ultra-violet is given by SWINGS (1942). Nevertheless, it would be of the greatest interest to carry out such a programme and test whether the distribution of energy in the ultra-violet continuum is consistent with the relative numbers of quanta emitted by ions of various ionization potentials. Owing to the confluence of numerous absorption lines near each series limit, the spectra will not in practice have the intense discontinuities predicted by models based on continuous absorption alone (STRUVE, 1951). (b) is probably a difficulty only in the case of hydrogen, since the limited extent of the shells corresponding to the higher ionization potentials in planetary nebulae and emission envelopes (see below) suggest

complete absorption of the ionizing radiation. (c) is a difficulty that can be overcome very readily in the case of hydrogen by computing the relative probabilities of recombination to various levels from the formula of CILLIÉ (1932); these formulae can probably be extended approximately to other ions. The correction for hydrogen is not large, since the observed lines $H\alpha - H\delta$ obtain about $\frac{1}{4}$ of the Lyman quanta (PAGEL, 1958) even if Lyman lines are able to escape, and multiplication of the number of quanta by this factor corresponds to an increase of only a few thousand degrees in the "photo-electric" temperature.

The self-absorption effect (d) can be examined by comparing the ratio of the intensities of two lines originating from the same upper level, e.g. a Paschen line and a Balmer line, with the ratio of the corresponding transition probabilities. Such measurements normally require spectrophotometry over a wide spectral region and infra-red spectra are thus particularly valuable. A study of the Paschen and Balmer emission lines (including the Balmer decrement) by E. M. and G. R. BURBIDGE (1955) showed that self absorption is probably absent in the hydrogen emission lines of P Cygni, but present in those of the pole-on Be star χ Ophiuchi. The exact effect of self-absorption depends on the solid angle subtended by the star at a typical point of the shell; in a large shell, scattered quanta still manage to leave eventually (though fluorescence effects will degrade them into fundamental lines of a series, such as $H\alpha$), but in a small one a substantial proportion will be absorbed by the central star and thus prevented from appearing in the emission spectrum.

Selective scattering and fluorescence (e) have already been discussed in Section 2. In general, their effect can only be determined by a detailed discussion of the entire observable spectrum of a given species (demonstrating the presence or absence of selectivities) and of the state of excitation and ionization. Generally speaking, the existence of a recombination spectrum of an ion A implies that A^+ is the predominant ion, so that A itself is not sufficiently abundant to cause much scattering unless the corresponding element is very abundant indeed (e.g. hydrogen) or unless the density is very high. Thus (e) probably represents a complete breakdown of the recombination hypothesis rather than merely a correction to it, unless there is a well-marked stratification into an A^+ shell, emitting the spectrum of A by recombination, and an A shell, absorbing this emission and re-radiating it through scattering and fluorescence. If there is a sufficient difference between the two shells in the velocity component along the line of sight, the emission and absorption processes may, to a first approximation, be considered in isolation from each other.

The consequences of BEALS' hypothesis for WR stars were explored in detail by ZANSTRA and WEENEN (1950), who pointed out that the regions of emission by successive ions must be distributed with increasing distance from the star according to the following scheme, as represented by the tabulation on the top of p. 220. One restriction is that an ion of an element of low abundance may continue to emit its spectrum (due to recombination of the next higher ion) out to the next boundary in the scheme, where all of its ionizing radiation will have been absorbed by the next ion in order of decreasing ionization potential. Also, there may only be enough material in the shell to absorb radiation above a certain frequency that is higher than the limit λ 912; in this case, the portion to the right of a certain vertical line will be cut off. Here ZANSTRA and WEENEN point out the following very interesting confirmation of the recombination hypothesis: in those cases where the WR star is the nucleus of a planetary nebula, the scheme carries on without discontinuity into the nebula


Absorbing ion
λ31·6 λ109 λ127 λ160 λ160 λ192 λ225 λ228 λ259 λ260 λ353 λ419 λ504 λ508 λ852 λ911 λ912
C4+ O4+ N4+ O3+ N3+ C3+ O2+ He+ C2+ N2+ O+ N+ He C+ N O H
Spectra emitted
O V O IV O III O II O I
N V N IV N III N II N I
C IV C III C II C I
He II He I
H

itself. This result, which is of course quite inconsistent with any approximation to local thermodynamic equilibrium, has been confirmed in more detail by VAN PELT (1957); it even holds for the forbidden lines and for lines excited by BOWEN's fluorescence mechanism, bearing in mind that these are both emitted by the "absorbing" ion, one stage lower than the "recombining" ion. In these cases, the division occurs between λ 192 and λ 350.

ZANSTRA and WEENEN proceed to analyze the equivalent widths of four WR stars measured by BEALS (1934). Using the continuum of the underlying star as a standard of intensity, the number of quanta emitted by one ion in the whole envelope is

$$\Sigma L' = \frac{10^5}{c} 4\pi^2 R_*^2 \Sigma\{w B_\nu(T_*)\}, \qquad \ldots . (1)$$

where w is the equivalent width measured in km/sec and T_* the stellar brightness temperature in the photographic range. Using CILLIÉ's theory, the rate of recombinations per cm^3 is

$$C_i(R) = N_i(R) N_e(R) Z.\, 3{\cdot}26 \times 10^{-6} \sum_{n=j}^{\infty} M\left(n, \frac{T_e}{Z^2}\right), \qquad \ldots . (2)$$

where T_e is the electron temperature at a distance R from the star and $M\ (n, T_e)$ is tabulated by CILLIÉ (1932). Z is the charge on the "recombining" ion and j the principal quantum number of the lowest upper level from which observable quanta are emitted; infra-red quanta are neglected. Assuming the ionization scheme to be cut off (Model II) between λ 228 (He^+) and λ 259 (C^{2+}), one has three regions: region 1 with C IV and He II emissions; region 2 with C III and He II; and region 3 (extending to infinity) with C III and He I. By equating (1) and (2) for each region and summing the contributions to He II and C III emissions from the pairs of regions 1, 2, and 2, 3 respectively, 5 equations are obtained from which the abundance ratio N_c/N_{He} and the four other unknowns $\Sigma_{1,2} L'_{He\ II}$ and $\Sigma_{2,3} L'_{C\ III}$ are easily derived. The result, which is virtually independent of T_* and T_e over a wide range, indicates

a ratio C : He of 0·5 in two cases and 0·2 in a third; these three stars belong to the WC sequence. A similar analysis of the He : H ratio in a fourth WR star HD 192163 leads to the conclusion that hydrogen and helium are about equally abundant, in agreement with the conclusion reached by Neubauer and Aller (1948) for HD 45166, on the assumption of thermal equilibrium. Since the analysis is independent of any assumptions regarding the distribution of ionizing radiation, it seems difficult to escape the conclusion that these abnormally high C abundances are real, though Miss Underhill (1958b) has challenged the assumption of complete ionization in each shell. Miss Underhill makes out a plausible case for attributing the bifurcation into WC and WN sequences to the characteristic shapes of the stellar ultra-violet continuum at different effective temperatures, the C spectra corresponding to about 50,000° and the N spectra to about 62,000°. However, her analysis ignores the effect of absorption lines (which would tend to blur the discontinuities in the continuous spectrum) and of ionization from excited levels (which becomes important when the dilution factor and the optical depth of the shell for resonance lines are high); furthermore, line widths and the general state of excitation vary similarly along parallel sequences in the WC and WN stars (Beals, 1940). Finally, Zanstra and Weenen's C : He ratio is confirmed in a separate analysis on the basis of an ionization temperature of 80,000° by Weenen (1950), who finds this approach to be more self-consistent than the pseudo-thermodynamic excitation equilibrium assumed by Aller (1943); Weenen's results are fairly insensitive to the ionization temperature adopted, since the relevant Boltzmann factors involve only the small ionization potentials from highly excited levels, but errors by a factor of 2 or so could easily arise. The remainder of Zanstra and Weenen's paper will be described in the next subsection.

The results of Zanstra and Weenen, and of van Pelt, provide impressive qualitative confirmation of Beals' hypothesis although there are many indications that the dilution factor is large, of the order of 0·1. This provides somewhat of a theoretical puzzle, since calculations by Miyamoto (1952) and Kogure (1953) indicate that ionization is mainly from excited levels (so that Beals' photoelectric temperatures near 80,000° are too high by a factor 2 or so) and that the Rosseland cycle is not strongly operative. However, observations of the eclipsing binary V444 Cygni tend to confirm Beals' figures (see next section). In the light of Sobolev's work, discussed below, it seems likely that the key to the puzzle lies in the expansion of the shell, which enables line radiation to escape through the "leakage" effect and thus acts in the same way as a low geometrical dilution factor.

(ii) *The Sizes and Densities of Wolf-Rayet Shells*

The diameters of WR stars are somewhat uncertain. Assuming effective temperatures in the neighbourhood of 80,000°, Beals (1940) finds a mean radius of 0·38 $R_\odot$ for the planetary nuclei and 2·2 $R_\odot$ for the ordinary WR stars. Beals infers from the smallness of these radii, compared with normal O stars of the same temperature, that the WR stars should have relatively small mass; otherwise the surface gravity would be too high to permit the ejection of material. No doubt the lower ejection velocities observed in planetary nuclei are correlated with their lower radii (Swings, 1942).

With regard to the size of the shell, the following features (Swings, 1942) suggest

that—in accordance with dynamical theory (CHANDRASEKHAR, 1934b)—the emitting region is not much bigger than the star itself ($R < 5R_*$ at the most):

1. The absence of dilution effects in the violet-displaced absorption components of He I.

2. The widths of the absorption components, assumed to arise from the difference between the velocity components along the line of sight of uniformly expanding material seen against the centre and edge of the stellar disk, suggest $R \simeq 1{\cdot}2\,R_*$ on the average. This again presents something of a puzzle, since the occultation effect should produce a distinct asymmetry in the emission-line profiles (CHANDRASEKHAR, 1934b), which is not observed.

Results of fundamental importance on this question have been obtained from studies of the double-lined spectroscopic and eclipsing binary V444 Cygni (HD 193576), which consists of an O-type star and a WN 5 star revolving in circular orbits with a period of 4^d2. The spectrographic orbit was analyzed by O. C. WILSON (1940), who found a minimum mass of $9{\cdot}7\odot$ for the WN component. The eclipsing binary character was demonstrated by S. GAPOSCHKIN (1941) and O. C. WILSON (1942) rediscussed the results with the aid of microphotometer tracings and a preliminary photoelectric light curve by KRON (KRON and GORDON, 1943), finding a radius of $18{\cdot}5\,R_\odot$ and a mass of $12{\cdot}4\odot$ for the WN component. WILSON argued further that, on BEALS' hypothesis of an expanding envelope, the emission lines should have the velocity of the underlying star at the moment when the emitting gases were ejected, leading to a phase lag in the measured velocity. Since the observations exclude a phase lag greater than 0·01 of the period, corresponding to a travel-distance of only 7×10^6 km $= 10R_\odot$, WILSON concluded that the emission bands must be shifted to the violet through occultation by some hundreds of km/sec. Since this is not observed (on the contrary, there is a constant red-shift, of unknown origin, of 90 km/sec), this observation cast serious doubt on the expansion hypothesis. BEALS, on the other hand, carried out a detailed spectro-photometric study (BEALS 1944), from which he concluded that "tidal" effects (already noted by WILSON for He II λ 4686) would remove the spherical symmetry and therefore upset the basic assumptions of WILSON's "transit-time" hypothesis. In contrast to WILSON, who found the WN component to be the brighter, BEALS deduced from a comparison of the absorption and emission lines with normal stars of the same spectral classes that the O star is 1^m7 brighter than the W star, in agreement with the original hypothesis of GAPOSCHKIN (1941). Assuming an effective temperature of 80,000°, he obtained a radius of only $1{\cdot}3\,R_\odot$; this temperature may be too high, but the radius is fairly insensitive to it (roughly, visual luminosity $\propto R_*{}^2T_*$). Also, BEALS noted that the photoelectric light curve obtained by KRON and GORDON shows the primary minimum (W star in front) to be twice as wide as the secondary minimum; this suggests that the W star is actually quite small, with a high effective temperature, and is surrounded by a large non-luminous electron shell (KOPAL, 1944). Later analysis of the scattering coefficient suggested a maximum electron density of 2×10^{12} cm^{-3} (KOPAL and M. S. SHAPLEY 1946). This value agrees well with the interpretation of the strength of He II λ 4686 on the recombination hypothesis (ALLER, 1954). Finally, BEALS found a violet-displaced absorption component of He I λ 4471 which reaches maximum strength at secondary minimum, when it appears to be absorbing light from both components; fainter red-displaced absorption components of He II and hydrogen were observed at primary minimum, presumably arising from gas between

the two components and absorbing the light of the O star only. Hence there is direct evidence of extensive shells showing stratification (He I round both stars and He II between them) in accordance with Beals' hypothesis.

Evidently V444 Cygni is a more complex system than had been assumed in the early analyses. Kron and Gordon (1950) rediscussed their light curve, centred on λ 4500, in conjunction with a new light curve in the red, centred on λ 7200. The shape of primary minimum was shown to be independent of wave-length, in accordance with Kopal's hypothesis of electron scattering in a highly ionized medium. Owing to distortion of the shape of primary minimum by this process, the geometrical analysis was carried out solely from the blue observations of secondary minimum, using the magnitude difference determined by Beals; a similar, though smaller, distortion in the secondary minimum (presence of wings) was attributed to the gradual eclipse by the O-star of non-negligible re-emission from the electron halo. Adapting the standard procedure for geometrical analysis of light curves to the case of a uniform circular stellar disk surrounded by a fainter limb-darkened scattering disk, the secondary minimum indicated radii of $2{\cdot}2\ R_\odot$ for the W core, $7\ R_\odot$ for the luminous disk, and $10\ R_\odot$ for the O star, the separation of the components being in the neighbourhood of $40\ R_\odot$. A similar hypothesis (with different radii) fits the primary minimum; a completely opaque core of radius $4{\cdot}4\ R_\odot$ is surrounded by a semi-transparent disk with transmission coefficient 0·86 and radius $17\ R_\odot$. The uniformity of this transmission coefficient suggests the presence of a thin spherical shell of ionized material surrounding the W star, a hitherto unsuspected result. Since the analysis is only concerned with projections in the plane normal to the line of sight, these disks could all equally well represent spheroids prolonged along the line joining the two components.

Hence the results of Kron and Gordon are, on the whole, consistent with Beals' analysis: a high effective temperature (though perhaps not as high as the 80,000° derived by Zanstra's method) is confirmed and the He II and He I shells may be identified with the opaque and the semi-transparent disks respectively. The luminous shell has a brightness temperature in the blue of 16,500°, corresponding to the measured colour temperatures of WR stars in general, and has about the same total luminosity as the bright core in this region; in the red, however, it is fainter than the core, so that its colour temperature turns out to be higher than that of the hot core. The absorption coefficient of the large spherical shell is equal in the blue and red, consistent with Thomson scattering. Some difficulties remain, however, in particular the absence of a reduction in emission line intensity during secondary eclipse; indeed, in the case of the eclipsing WR binary CQ Cephei, He II λ 4686 is actually enhanced during eclipses (Hiltner, 1950). These effects must be attributed to departures from spherical symmetry, probably correlated with the effects described by Beals. However, Hiltner's observations failed to demonstrate the presence of a significant amount of polarization in the continuum between eclipses.

Further difficulties are pointed out in a spectrographic and theoretical study of V444 Cygni by Münch (1950), whose observational results differed from those of O. C. Wilson in the velocities assigned to the γ-axes of the absorption and emission components; in agreement with previous results of Keeping (1947) for this star and of Struve (1944) and Hiltner (1945) for other WR stars, the latter vary from line to line (-40 km/sec for N IV λ 4057 and $+70$ km/sec for N V, mean of $\lambda\lambda$ 4603, 4619, the "true" γ-velocity derived from H 8, H 9 and H 10 in absorption being $+10$ km/sec).

MÜNCH also pointed out complex variations in the appearances of the emission lines and of the violet absorption edges of He I (previously noted in other cases by BEALS and HILTNER). Furthermore, the absorption lines H 9 and H 10 of the O star appear broad and shallow at primary eclipse, an effect explained by MÜNCH on the basis of non-coherent Thomson scattering in an equivalent plane layer of optical depth $\tau \simeq 0{\cdot}5$, part of which probably arises in the outer layers of the O star itself. Furthermore, an appreciable part of the luminosity of the scattering envelope is shown to arise from scattering (and probably fluorescent re-emission) of light from the O-type companion; the resulting reflection effect might explain why the envelope does not become perceptibly fainter when the O star passes in front of it.

MÜNCH points out some very interesting selectivities in the appearance of absorption edges to the emission lines of V444 Cygni, recalling the behaviour of Of stars. The N III multiplet $3p\,^2P^o - 3d\,^2D$, which is the only one to appear in emission in Of stars such as 9 Sagittae, here increases in intensity at secondary minimum much more than do other lines of the same ion. Similarly, the N IV multiplet $3p\,^1P^o - 3d\,^1D$, which appears in emission in 9 Sge (SWINGS and STRUVE 1940b), appears here as pure emission; while the $3s\,^3S - 3p\,^3P^o$ lines (absorption lines in 9 Sge) develop an absorption edge at primary minimum. These effects undoubtedly represent the influence of selective fluorescence excited by the radiation of the O star; hence MÜNCH's conclusion that they invalidate BEALS' hypothesis in general seems unjustified.

On the other hand, the presence of an electron-scattering envelope of appreciable optical depth indicates that the calculation of line profiles without including self-absorption effects is unjustified. THOMAS (1947, 1949) found that self-absorption in a uniformly expanding atmosphere led to a peak at the violet end of the emission band (similar to the occultation effect), which is not observed; he concluded that the simple expansion hypothesis must be incorrect, and proposed a model in which the density gradient found for V444 Cygni is supported by vertical turbulent motions whose amplitude decreases outwards. The excitation is qualitatively explained by statistical equilibrium with an electron temperature near 10^5, and with no stratification, a conclusion that is inconsistent with the results described in the last subsection. MÜNCH considers a compromise model in which expansion occurs with an outward deceleration, with turbulence, and in which the excitation increases outwards (i.e. towards the O star). Under these conditions, the electron optical depth is sufficient to broaden the line profiles considerably, although expansion still plays the major part. The observed symmetry of the emission profile is attributed to the washing out of irregularities by continuous scattering and by the violet-shifted absorption edge (which, owing to deceleration, now penetrates nearer the centre of the line). This absorption edge, especially when intensified by its simultaneous action on both continua at times of conjunction, explains the anomalous red shift; this is absent in the N IV line λ 4603 because of the absorption edge of another line on its red side. The same explanation of the red shift was given earlier by SOBOLEV (1947).

While MÜNCH's picture removes some of the difficulties inherent in BEALS' hypothesis, and may indeed fit the case of V444 Cygni and similar systems, the outward increase of excitation cannot be true in general in view of the results of ZANSTRA and WEENEN and VAN PELT. In view of the scrambling effects of electron scattering on the emission-line profiles, the difficulties due to the absence of noticeable effects of occultation or self-absorption do not appear fatal, though an exact treatment similar to that of ROTTENBERG, taking the curvature of the atmosphere into account,

would be of great interest. O. C. Wilson's transit-time difficulty rests on dynamical assumptions that may not be valid. It therefore appears that Beals' hypothesis is still a fruitful one, and indeed one that is confirmed by several independent observations.

With this in mind, we may return to Zanstra and Weenen's paper, in the second part of which the radii of the emitting shells are calculated. Assuming uniform expansion, continuity requires the ion density of the dominant ion at distance R to be

$$N_i(R) = N_i(R_0)\frac{R_0^2}{R^2} = N_A\frac{R_0^2}{R^2}, \quad \ldots\ldots (3)$$

while the corresponding electron density in the sth shell is

$$N_{e,s} = \sum_s ZN_i(R) = \frac{R_0^2}{R^2}\sum_s ZN_A = \frac{R_0^2}{R^2}q_s N_{\mathrm{E}}(R_0), \quad \ldots\ldots (4)$$

where

$$q_s = \frac{\sum_1^s ZN_A}{\sum_1^\infty ZN_A}. \quad \ldots\ldots (5)$$

In Model II,

$$q_1 = 1,$$

$$q_2 = (3N_{\mathrm{C}} + 2N_{\mathrm{He}})/(4N_{\mathrm{C}} + 2N_{\mathrm{He}}),$$

and

$$q_3 = (3N_{\mathrm{C}} + N_{\mathrm{He}})/(4N_{\mathrm{C}} + 2N_{\mathrm{He}}),$$

so that the q's are known from the relative abundance of C and He if electrons supplied by hydrogen and other elements are neglected. Using these results, one can now integrate (2) over the volume of each shell and equate to (1):

$$\sum_s L' = \int_{R_{s-1}}^{R_s} 4\pi R^2 C_i(R)\,dR = \\ = N_A N_e(R_0) 4\pi \cdot 3{\cdot}26 \times 10^{-6} Z \sum_j^\infty M\left(n, \frac{T_e}{Z^2}\right) R_0^4 q_s \left(\frac{1}{R_{s-1}} - \frac{1}{R_s}\right). \quad \ldots\ldots (6)$$

$\Sigma L'$ is known from the line intensities and the previous analysis. $R_3 = \infty$, and the ratios R_1/R_0, R_2/R_0 are now determined from two simultaneous equations involving the observed quantum ratios C IV : C III and He II : He I. Assuming $T_* = 50{,}000°$ (the photo-electric temperature obtained from hydrogen), $R_0 = R_*$, it is now possible to derive the absolute densities of electrons and of each ion from the equivalent widths of the emission lines; assuming $R_* = R_\odot$ one obtains $N_e(R_0) = 10^{13}$, in fairly good agreement with other determinations. $R_1/R_0 \simeq 1{\cdot}1$, while R_2/R_0 ranges from 1·1 to 1·6, in agreement with other evidence for a small emitting shell. Finally, these figures are used to verify the assumption of complete ionization in each shell.

(iii) *The Sizes and Densities of Transparent Envelopes from Hydrogen Line Intensities: P Cygni Stars*

Assuming Beals' hypothesis, it is possible to apply an analysis similar to that of Zanstra and Weenen to determine the densities and equivalent radii of transparent

shells of P Cygni stars, in which the spectrum of hydrogen is prominent. The evidence from the Paschen: Balmer ratio and from the "nebular" decrement that self-absorption is negligible in the Balmer spectrum of P Cygni itself (BURBIDGE and BURBIDGE, 1955) has already been mentioned.

We assume the envelope to be opaque in the Lyman series, so that each Lyman line emitted is absorbed at another point (ZANSTRA, 1926; Case B of BAKER and MENZEL, 1938); in this case, each downward transition to the ground level is balanced by an inverse transition, so that the existence of this level may be ignored. The number of quanta emitted per cm^3 per sec in a Balmer line is simply

$$Q_{\mathrm{H}_n} = N_n A_{n2}, \qquad \ldots . (7)$$

where N_n is the number density of hydrogen atoms in the nth level and $A_{nn'}$ is the relevant transition probability ($n' < n < n''$). Neglecting photo-excitation by line radiation and photo-ionization from the third and higher levels, the steady-state condition gives the equations of statistical equilibrium for each level:

$$N_n A_{n2} + N_n \sum_{n'=3}^{n-1} A_{nn'} = C_n + \sum_{n''=n+1}^{\infty} N_{n''} A_{n''n}, \qquad \ldots . (8)$$

where C_n is the rate of spontaneous recombination to the nth level, calculated by CILLIÉ (1932). Stimulated recombinations are virtually balanced by photo-ionizations at high temperatures. Summing (8) over $n = 3$ to ∞, we have

$$\sum_{n=3}^{\infty} N_n A_{n2} + \sum_{n=3}^{\infty} N_n \sum_{n'=3}^{n-1} A_{nn'} = \sum_{n=3}^{\infty} C_n + \sum_{n=3}^{\infty} \sum_{n''=n+1}^{\infty} N_{n''} A_{n''n}, \qquad \ldots . (9)$$

in which the double sums on the two sides of (9) are equal (cf. MENZEL, ALLER and BAKER, 1938). Hence

$$\sum_{n=3}^{\infty} N_n A_{n2} = \sum_{n=3}^{\infty} C_n, \qquad \ldots . (10)$$

i.e. the total number of quanta emitted in the Balmer series is equal to the total number of recombinations on levels $n \geqslant 3$, a result which is reasonable on general grounds since each recombination must ultimately lead to the emission of a Balmer quantum and other sources of excitation are neglected. At some, probably fairly high, value of n our assumption of opacity to Lyman lines must break down, requiring the application to each term on the left of (10) of a correction factor $(A_{n1} + A_{n2})/A_{n2}$. However, since 90 per cent of the Balmer quanta are in fact emitted in H$\alpha - \epsilon$, this correction is probably of little importance in most cases.

(10) may now be applied to measurements of equivalent width. Assuming the star to radiate in the observable region like a black body of temperature T_* and neglecting occultation, we have

$$4\pi \int_{R_*}^{\infty} \sum_{n=3}^{\infty} C_n(R) R^2 dR = 8\pi^2 c R_*^2 \sum_{n=3}^{\infty} \frac{w_n}{\lambda_n^4 (e^{h\nu_n/kT_*} - 1)}, \qquad \ldots . (11)$$

where w_n is now the equivalent width in cm. Furthermore,

$$C_n(R) = N_e(R) N_p(R)\, 3{\cdot}26 \times 10^{-6}\, M(n, T_e) \qquad \ldots . (12)$$

and, assuming uniform expansion from the stellar surface,

$$N_p(R) = qN_e(R) = qN_e(R_*)R_*^2/R^2, \quad \ldots . (13)$$

where $q < 1$.

Combining (11), (12), and (13),

$$qN_e^2 = 8{\cdot}3 \times 10^5 \left(\frac{R_*}{R_\odot}\right)^{-1} \frac{\sum_{n=3}^{\infty} \frac{w_n}{\lambda_n^4 (e^{h\nu_n/kT_*} - 1)}}{\sum_{n=3}^{\infty} M(n, T_e)}, \quad \ldots . (14)$$

where $N_e(R_*)$ has been replaced by N_e.

We confine our attention to two cases: P Cygni itself, for which measurements have been carried out by BEALS (1950), BURBIDGE and BURBIDGE (1955), and others; and AG Pegasi, which is of particular interest owing to the presence of Zeeman splitting (BABCOCK, 1958), and has recently been studied by BEALS (1950), MERRILL (1951a, b), by TCHENG MAO-LIN and BLOCH (1952) and by BURBIDGE and BURBIDGE (1954). In general, we take the BURBIDGES' measurements. However, their measured equivalent width of Hα in P Cygni appears to be too low, owing to over-exposure, and the value given was increased by 40 per cent, to bring it up to the value expected from the Balmer decrement. In AG Pegasi Hα was not measured, and a hypothetical value was deduced from the Balmer decrement; Hβ and Hγ were increased by about 50 per cent to allow for an apparent self-absorption effect (which may also be due to over-exposure). These points are of some importance, because H$\alpha - \gamma$ contribute the vast majority of the Balmer quanta. The details are given in the following Table I.

Table I

	P Cyg	AG Peg (HD 207757)
T_* (assumed)	20,000°	25,000°
T_e (assumed)	15,000°	20,000°
$\sum \frac{w_n}{\lambda_n^4 (e^{h\nu/kT_*} - 1)}$	$3{\cdot}2 \times 10^{10}$	$1{\cdot}3 \times 10^{12}$
$\sum M\ (n, T_e)$	$3{\cdot}7 \times 10^{-8}$	$2{\cdot}7 \times 10^{-8}$
$N_e q^{\frac{1}{2}}$	$8{\cdot}5 \times 10^{11}\ (R_*/R_\odot)^{-\frac{1}{2}}$	$6{\cdot}4 \times 10^{12}\ (R_*/R_\odot)^{-\frac{1}{2}}$

In Section 1, the condition that the tangential optical depth due to electron scattering be unity at the limb led to the result

$$N_e = 6{\cdot}85 \times 10^{12} (R_*/R_\odot)^{-1}. \quad \ldots . (15)$$

Combining this with the results obtained above, assuming $q = 1$, we obtain the following solutions for R_* and N_e; Table II.

The resulting values of R_* may be compared with those derived from the absolute magnitudes given by BEALS (1950) and from our assumed effective temperatures T_*, using the equation

$$\log(R_*/R_\odot) = \frac{5810}{T_*} - 0{\cdot}20 M_{\text{vis}} - 0{\cdot}04 + 0{\cdot}5 \log(1 - 10^{-11{,}570/T_*}) \quad \ldots . (16)$$

(ALLEN, 1955). The resulting radii differ by factors of the order of 2, which is not surprising in view of our rather arbitrary choice of T_*, T_e and also in view of the uncertainty in $M_{\rm vis}$. For example, Miss ROMAN (1951) has attributed the observed colour of P Cygni entirely to interstellar reddening, leading to a corrected absolute magnitude $M_{\rm vis} = -8{\cdot}3$, which would give a radius in fair agreement with our high value. On the other hand, our value of $1{\cdot}15\ R_\odot$ for AG Peg seems less likely than the value of $3{\cdot}8\ R_\odot$ deduced from the absolute magnitude. Possibly equation (15) is inapplicable owing to an appreciable contribution of neutral hydrogen to the continuous opacity. Owing to the uncertainties involved (in T_*, T_e, and $M_{\rm vis}$), it does not seem worth while to try to improve the agreement by adjusting the unknown

Table II

	P Cyg	AG Peg
$R_*/R_\odot$ from (14), (15)	65	1·15
Ne from (14), (15)	$1{\cdot}0\times10^{11}$	$6{\cdot}0\times10^{12}$
$M_{\rm vis}$ (Beals 1950)	$-6\overset{m}{.}3$	$-2\overset{m}{.}4$
$R_*/R_\odot$ from (16)	28	3·8
N_e from (14)	$1{\cdot}6\times10^{11}$	$3{\cdot}3\times10^{12}$

parameters. However, we do end up with fairly reliable values of N_e, which should lie somewhere between the two values given in the table. It is of interest to note that the mean value of N_e for AG Peg deduced by the BURBIDGES from the INGLIS-TELLER formula is $1{\cdot}5\times10^{12}$; it is quite reasonable that this should be slightly less than our value, since it refers to the emitting region as a whole (see equation (20)), and the order-of-magnitude agreement seems to be strong evidence that the emission occurs in the immediate neighbourhood of the surface. This result was reached by the writer by a somewhat different, though logically equivalent, method (PAGEL, 1958), in contrast to the BURBIDGES' conclusion that the shell extends to 400 stellar radii. In the same paper, the writer obtained an estimate of $4{\cdot}7\times10^{11}$ for the mean electron density in the P Cygni envelope; according to the present results, this value appears to be much too high, probably because the assumption of complete absorption of Lyman continuum radiation, on which it was based, does not hold.

This last point is of some interest, and we now test it by means of PANNEKOEK's ionization equation:

$$\frac{N_1(R)}{qN_e^2(R)} = \frac{1}{2W}\left(\frac{h^2}{2\pi mkT_1}\right)^{\frac{3}{2}}\left(\frac{T_1}{T_e}\right)^{\frac{1}{2}} e^{X_1/kT_1}$$

$$= \frac{2R^2}{R_*^2}10^{-21{\cdot}383+68538/T_1}\left(\frac{T_1}{10^4}\right)^{-1}\left(\frac{T_e}{10^4}\right)^{-\frac{1}{2}}, \qquad \ldots\ldots(17)$$

where T_1 is the brightness temperature beyond the Lyman limit, here taken as equal to T_*. The radial optical depth at the Lyman limit is

$$\tau_1 = \alpha\int_{R_*}^{\infty} N_1(R)dR$$

$$= 10^{-27{\cdot}44}\left(\frac{R_*}{R_\odot}\right)qN_e^2\left(\frac{T_1}{10^4}\right)^{-1}\left(\frac{T_e}{10^4}\right)^{-\frac{1}{2}}10^{68538/T_1}. \qquad \ldots\ldots(18)$$

Taking $(R_*/R_-)qN_e^2$ and the temperatures from Table I,

$$\tau_1 = 0{\cdot}29(\text{P Cyg}) \quad \text{and} \quad 2{\cdot}31\ (\text{AG Peg}).$$

We thus see that, with the figures given, the envelope of P Cygni is completely ionized at all distances, while that of AG Pegasi is a bounded H II region (since $\tau_1 > 1$). This fact is perhaps correlated with the rather curious observation of TiO bands in the spectrum of AG Peg, which has been attributed to the presence of a late-type companion (BEALS, 1950). In our view, it seems not unreasonable that a late-type spectrum might form in the outlying H I region, where high-frequency radiation is no longer present; however, detailed calculations will be needed to test this possibility. This theory has been put forward by SOBOLEV (1947) to account for late-type stars with emission lines.

The H II region comes to an end at a distance R_1 from the star, where

$$\tau_1(R_1) = 1 = \tau_1(1 - R_*/R_1) = 2{\cdot}31(1 - R_*/R_1). \qquad \ldots . (19)$$

Hence $R_1 = 1{\cdot}77\ R_*$. This in turn requires a correction to the result in Table I, since this was derived on the assumption that the emitting region extends to infinity. However, R_1 has been under-estimated here because the dilution factor near the surface is greater (nearly by a factor 2) than the usual approximation $R_*^2/4R^2$ assumed here. Hence τ_1 will actually be of the order of 1·5 and $R_1 \sim 3R_*$. Furthermore, we have neglected ionization from the second quantum level. In view of the uncertainties, it is not worth while to carry the analysis further at this point.

Before leaving the uniform-expansion model, it is of interest to calculate the mean electron density $\bar{N}_e$ and the mean dilution factor $\bar{W}$ in the shell, assuming this to extend to infinity. Since the rate of emission per unit volume at radius R is proportional to $N_e^2(R)$, we have

$$\bar{N}_e = \frac{\int_{R_*}^{\infty} N_e^3(R)R^2dR}{\int_{R_*}^{\infty} N_e^2(R)R^2dR} = \tfrac{1}{3}N_e. \qquad \ldots . (20)$$

Similarly,

$$\bar{W} = 1/12 \qquad \ldots . (21)$$

from the simple formula for W, though this is actually a lower limit. The value $\bar{N}_e$ obtained from (20) for AG Peg is in excellent agreement with the BURBIDGES' value derived from the Inglis-Teller formula. The observational measurement of W is more difficult, even if it is assumed that the absorption lines arise in the same region as the emission lines. STRUVE and ROACH (1939) obtained $W \simeq 0{\cdot}05$ for He I and $W > 1$ for H, which they suggested might be due to metastability of the 2s level. BURBIDGE and BURBIDGE (1955) overcame this difficulty by studying the Paschen absorption lines and obtained $W \simeq 0{\cdot}5$. The values for He and H are thus rather discrepant, but they bracket the theoretical value of $\bar{W}$ quite satisfactorily.

We thus see that the idealized picture of uniform expansion gives a fairly satisfactory representation of the observed intensities of lines in P Cygni stars. It is also very attractive mathematically. However, there must be considerable turbulent

motions present; the picture of an optically thin, uniformly expanding envelope is inconsistent with the observed line profiles. Hence it is of interest to analyze the emission intensities on the basis of an envelope with uniform electron density $\bar{N}_e$, a model which is just as simple and may represent a crude approximation to a situation with strong turbulence. In this case the emitting envelope is regarded as roughly spherical, with a sharp boundary at a distance aR_* from the centre of the star (BURBIDGE and BURBIDGE, 1954, 1955). The method used by the BURBIDGES to derive a from the hydrogen emission-line intensities has been criticized in detail and an alternative analysis proposed (PAGEL 1958); this analysis assumed that the populations of the higher levels are given by the statistical-equilibrium calculations for Case B in a planetary nebula (BAKER and MENZEL, 1938), which is logically equivalent to equation (10). We now adapt the discussion just given to the uniform-density model; the validity of the assumption will be discussed in the next sub-section.

Assuming the star to occupy or occult a cylinder of radius R_* and height $(a+1)R_*$, we have

$$\frac{4}{3}\pi R_*{}^3\left[a^3-\frac{3}{4}\left(a+1\right)\right]\sum_{n=3}^{\infty}C_n = 8\pi^2 c R_*{}^2\sum_{n=3}^{\infty}\frac{w_n}{\lambda_n{}^4(e^{h\nu_n/kT_*}-1)}$$

or

$$a^3-\tfrac{3}{4}(a+1) = 2{\cdot}49\times10^6\left(\frac{R_*}{R_-}\right)^{-1}\frac{1}{q\bar{N}_e{}^2}\frac{\displaystyle\sum_{n=3}^{\infty}\frac{w_n}{\lambda_n{}^4(e^{h\nu_n/kT_*}-1)}}{\displaystyle\sum_{n=3}^{\infty}M(n,\,T_e)}. \qquad \ldots\ldots(22)$$

The assumed values of $q\bar{N}_e{}^2$ and R_*/R_-, together with the resulting values of a, are given in Table III.

Table III

	P Cygni	AG Pegasi
R_*/R_-	65	3·8
$q^{\frac{1}{2}}\bar{N}_e$	$3{\cdot}3\times10^{10}$	$1{\cdot}1\times10^{12}$
a (this analysis)	3·2	3·1
a (BURBIDGE and BURBIDGE)	1·5	400
a (PAGEL 1958)	1·15	3·3

For AG Peg, the present value is entirely in agreement with that derived previously; for P Cygni it is much greater because of the lower electron density (by a factor 10) assumed here. The discrepancy from the BURBIDGES' value is due to their method of analysis, which involved the assumption that 1 per cent of the neutral hydrogen atoms are in upper levels giving rise to measurable Balmer lines.

The value $a = 3$ is actually inherent in our present analysis, since it is based on an electron density deduced from the uniform-expansion model; such a model is *always* equivalent to a uniform-density envelope of radius $3R_*$, as can be seen by comparing (14) with (22), bearing in mind (20).

(iv) *Statistical Equilibrium in a Transparent Hydrogen Envelope*

For a discussion of the Balmer decrement, it is necessary to know the distribution of hydrogen atoms among their various excited levels, including the continuum. If we write the combined BOLTZMANN-SAHA equation in the form suggested by MENZEL (1937), the density of hydrogen in the nth level is

$$N_n = qN_e^2 n^2 b_n \left(\frac{h^2}{2\pi mkT_e}\right)^{\frac{3}{2}} e^{\chi_n/kT_e}, \qquad \ldots\ldots (23)$$

where b_n is a numerical factor to be determined from the equations of statistical equilibrium for each level:

$$\sum_{n''=n+1}^{\infty} F_{n''n} + \int_{\nu_n}^{\infty} F_{\kappa n} d\nu + \sum_{n'=1}^{n-1} F_{n'n} = \sum_{n''=n+1}^{\infty} F_{nn''} + \int_{\nu_n}^{\infty} F_{n\kappa} d\nu + \sum_{n'=2}^{n-1} F_{nn'}, \qquad \ldots\ldots (24)$$

where $F_{n''n}$ represents a spontaneous transition from level n'' to level n and $F_{\kappa n}$, $F_{n\kappa}$ represent spontaneous recombination and photo-ionization respectively. In the previous subsection, stellar radiation was ignored (except as a source of photo-ionization from the ground level). This is the "nebular" case with $F_{nn''} = F_{n\kappa} = F_{n'n} = 0$; the appropriate values of b_n for $3 \leqslant n \leqslant 30$ have been calculated for various electron temperatures by BAKER and MENZEL (1938).

We now proceed one step further and inquire to what extent these results require modification in an envelope surrounding a hot star, assuming this to radiate as a black body of temperature T_*. Essentially, we have to include ROTTENBERG's scattering effect, together with fluorescence processes. The b_n's have been calculated for the nebular case C (Case A with inclusion of the stellar radiation field in the Lyman lines) by BAKER, MENZEL and ALLER (1938), but the very low values of W in nebulae permitted these authors to neglect induced emissions, which would be inappropriate here ($W \simeq 0{\cdot}1$). As has already been mentioned, the effect of photo-ionization and photo-excitation on the higher levels is greatly reduced in our case by the induced emissions, except in the case of photo-excitation from the second level; this suggests that the "nebular" b_n's (without the radiation field) may for high n be multiplied by a correction factor $(1+F_{2n}/2C_n)$, but that more complicated corrections will be required for the lower levels. The factor 2 in the expression $(1+F_{2n}/2C_n)$ arises from the fact that recombination and population by cascade from higher levels are about equally frequent for high n.

The various terms in (24) are given by the following expressions:

$$F_{n''n} = N_{n''} A_{n''n}, \qquad \ldots\ldots (25)$$

$$\int_{\nu_n}^{\infty} F_{\kappa n} d\nu = C_n = 3{\cdot}26 \times 10^{-6} qN_e^2 \bar{g} M(n, T_e), \qquad \ldots\ldots (26)$$

where $\bar{g}$ is the mean Gaunt factor for the relevant continuum.

$$F_{n'n} = WN_{n'} \frac{8\pi^2 e^2}{mc^2} f_{n'n} \frac{\nu_{n'n}^2}{e^{h\nu_{nn'}/kT_*} - 1} \left(1 - \frac{b_n}{b_{n'}} e^{h\nu_{nn'}/kT}\right). \qquad \ldots\ldots (27)$$

$$\int_{\nu_n}^{\infty} F_{n\kappa} d\nu = WN_n \frac{8\pi}{c^2} \alpha_n \nu_n^3 \int_{x_0}^{\infty} \frac{1 - \frac{1}{b_n} e^{-xT_*/T_e}}{x(e^x - 1)} dx, \qquad \ldots\ldots (28)$$

where $\varkappa_n$ is the absorption cross-section for photo-ionization at the absorption edge ν_n and $x_0 = h\nu_n/kT_*$. (28) may be rewritten in the approximate form

$$\int_{\nu_n}^{\infty} F_{n\kappa}\,d\nu \simeq WN_n\frac{8\pi}{c^2}\varkappa_n\nu_n^3\left(1-\frac{1}{b_n}e^{-x_0T_*/T_e}\right)E_1(x_0). \qquad \ldots\ldots (29)$$

For small x_0, (29) is invalid, but the whole term is then very small in any case.

A complete solution of equations (24)–(28) would involve a rather elaborate series of computations, suitable for an electronic computer. At present we restrict ourselves to the special case $T_* = T_e = 20{,}000°$, Case B, with the simplifying assumption that b_5 and higher b_n's may be obtained from the values calculated by Baker and Menzel (1938) by multiplication with the crude correcting factor $(1+F_{2n}/2C_n)$.

We then have the following equations of statistical equilibrium:

$$F_{24}+F_{34}+C_4+\sum_{n=5}^{\infty}F_{n4} = F_{42}+F_{43}+F_{4\kappa}+\sum_{n=5}^{\infty}F_{4n}, \qquad \ldots\ldots (30)$$

$$F_{23}+F_{43}+C_3+\sum_{n=5}^{\infty}F_{n3} = F_{32}+F_{3\kappa}+\sum_{n=5}^{\infty}F_{3n}, \qquad \ldots\ldots (31)$$

$$F_{42}+F_{32}+C_2+\sum_{n=5}^{\infty}F_{n2} = F_{2\kappa}+\sum_{n=5}^{\infty}F_{2n}, \qquad \ldots\ldots (32)$$

which must be solved simultaneously to yield b_4, b_3, b_2 as functions of W. The results are given in Table IV.

Table IV

Approximate values of b_n, including stellar radiation field, with $T_* = T_e = 20{,}000°$

	$W = 1{\cdot}0$	$W = 0{\cdot}5$	$W = 0{\cdot}1$	$W = 0{\cdot}05$	$W = 0{\cdot}01$	$W = 0{\cdot}0$
Wb_2	0·92	1·05	1·64	1·84	2·08	2·50
b_3	0·88	1·32	2·83	3·33	3·93	4·71
b_4	0·87	1·10	1·71	1·86	2·01	2·31
b_5	0·80	0·77	0·99	1·06	1·14	1·28
b_{10}	0·73	0·76	0·85	0·88	0·92	0·98

The b_n's for $W = 1{\cdot}0$ should of course all be unity and their departures from this value give a rough indication of the accuracy of the computation: b_2, b_3 and b_4 are good within about 10 per cent, while the higher b_n's become progressively worse. Thus our crude correction to the higher b_n's, though not sufficient to determine them accurately, does appear good enough to determine b_2, b_3, b_4. The values for $W = 0$ are quite similar to the results obtained by Baker, Menzel and Aller (1938) for Case C, when it is borne in mind that in our case the second quantum level is effectively the ground level: the b_n's are all greater than or equal to 1 and diminish gradually with increasing n, leading to a steeper Balmer decrement than the "nebular" one. For $n \geqslant 5$, b_n is never very far from 1, and $b_n = 1$ seems to be a suitable working approximation.

In Fig. 2 we compare the Balmer and Paschen decrements (in terms of numbers of quanta, taking $H\delta = 1$) predicted from Table IV ($W = 0{\cdot}1$) and from the "nebular" b_n's (Case B) with the measured values in the spectrum of P Cygni (BURBIDGE and BURBIDGE, 1955). Fig. 2 also shows the Balmer decrement for AG Pegasi (BURBIDGE and BURBIDGE, 1954). It is clear that the observed values agree better with the "nebular" decrement than with the present theory; this result is quite insensitive to our choice of T_* and T_e, although the relative level of the Balmer and Paschen lines is sensitive to T_*.

Since our analysis differs from that of BAKER and MENZEL (1938) mainly in including excitation by radiation in the Balmer lines, it appears that, in the envelopes of P Cygni and AG Pegasi, such excitation does not occur; its absence must be due

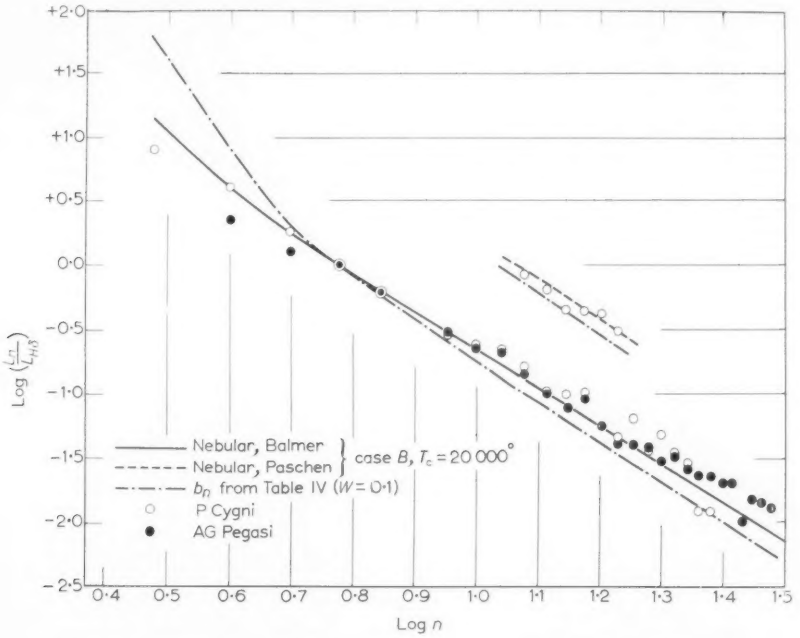


FIG. 2. Balmer and Paschen decrements in P Cygni, and Balmer decrements in AG Pegasi, both by numbers of quanta ($H\delta = 1$).

to the presence of strong absorption lines in the stellar spectrum as seen from a typical point in the envelope. Hence the Balmer absorption lines that are observed in P Cygni cannot be formed further out than the emission lines. They must in fact be formed near enough to the star for most of the re-emission to be removed by the occultation effect. The spectrum of AG Pegasi, which has varied considerably over the years, showed no hydrogen absorption lines when observed by the BURBIDGES in 1953; presumably underlying absorption lines are filled in by the very intense emissions. Owing to the DOPPLER shift caused by the expansion of the shell self-absorption appears to be absent except possibly in the very lowest members of the series (for which the measurements are uncertain) and the excitation of hydrogen is identical with that in a planetary nebula in which the radiation from the central star is negligible.

We conclude this discussion by calculating the optical depths of the envelopes of P Cygni and AG Pegasi to Balmer line radiation emerging from the central star, assuming no radial velocity-gradient and a Doppler broadening parameter of 50 km/sec. The optical depth is

$$\tau_{2n} = \frac{\pi^{\frac{1}{2}}e^2}{mc}\frac{f_{2n}}{\Delta\nu_{D,2n}}\int_{R_*}^{\infty} N_2(R)dR =$$

$$= \frac{\pi^{\frac{1}{2}}e^2}{mc}\frac{f_{2n}}{\Delta\nu_{D,2n}} 4\left(\frac{h^2}{2\pi mkT_e}\right)^{\frac{3}{2}} e^{\chi_2/kT_e}\int_{R_*}^{\infty} b_2\, qN_e^2(R)dR. \qquad \ldots\ (33)$$

Using (32) with $F_{2n} = 0$, and the higher b_n's calculated by BAKER and MENZEL for $T_e = 20{,}000°$,

$$b_2 W = 2{\cdot}79. \qquad \ldots\ (34)$$

Hence

$$\tau_{2n} = 0{\cdot}98 \times 10^{-21} qN_e^2(R_*/R_\odot)\lambda_{2n} f_{2n}, \qquad \ldots\ (35)$$

where λ_{2n} is in microns. Taking $qN_e^2(R_*/R_\odot)$ from Table I, we obtain the values of τ_{2n} given below (Table V).

Table V

Radial optical depths in Balmer lines

n	3	4	5	6	25
	Hα	Hβ	Hγ	Hδ	H25
τ_{2n} (P Cyg)	296	40·5	13·7	6·42	0·058
τ_{2n} (AG Peg)	$1{\cdot}68 \times 10^4$	2300	780	364	3·27

The values obtained for P Cygni are consistent with the observed "shell" absorption spectrum; they suggest that the emission from the lower members of the series should be cut down by self-absorption, but that higher members should be free from it (even when $\tau_{2n} > 1$) owing to the Doppler shift of the absorption profile along any ray that is not travelling radially through the expanding shell ("leakage" effect). We conclude that our analysis of the P Cygni envelope in terms of pure recombination without self-absorption has been substantially correct.

The optical depths for AG Pegasi, however, turn out to be so high as to cast serious doubt on our analysis, even though they should be cut down by a factor of 2·3 or so to take account of the fact that the ionized region extends to 1·8 R_* and not to infinity. With such high values of τ_{2n}, self-absorption must be important not only in the Balmer series, but probably in higher series as well. Under these conditions, the observed Balmer decrement is not very different from the "nebular" decrement, as has been shown by SOBOLEV (1947). SOBOLEV's theory will be discussed in Section 4(ii).

(v) *The Distribution of Hydrogen Atoms among Sub-levels*

The foregoing analysis has assumed that, for each value of the principal quantum number n, the relative populations of sub-levels with different azimuthal quantum numbers l are simply proportional to their statistical weights $(2l+1)$. The use of integrated transition probabilities $A_{nn'}$, rather than $A_{n\ l;\ n',\ l'}$, in subsection (iv)

was dependent on this assumption; on the other hand, our analysis in (iii), which simply equated the number of Balmer quanta to the number of recombinations, was not.

The assumption will not, in general, be true in the presence of purely radiative processes, since different sub-levels are formed by recombination at rates proportional to their statistical weights, but have different lifetimes. On the other hand, an equilibrium distribution among sub-levels will be brought about by collisions, if these are sufficiently frequent in comparison with radiative processes.

The rate of collisional excitation of $2s-2p$ transitions has been calculated by SEATON (1955). The predominant effect is due to proton rather than electron impact and the total transition rate is $4{\cdot}7\times10^{-4}N_e$ cm^3 sec^{-1} for $2s\rightarrow 2p$ and $1{\cdot}6\times10^{-4}$ N_e cm^3 sec^{-1} for the reverse process, assuming $T_e = 20{,}000°$, $q = 1$. Since the $2s$ level is metastable and the $2p\rightarrow 1s$ transition is assumed to be exactly balanced by its contrary, these processes compete with photo-ionization, the rate of which is $1{\cdot}2\times10^7W$ sec^{-1} if $T_*=20{,}000°$, and with population by cascade from higher levels, the total rate of which is of the order of $10^{-13}\,N_e{}^2/N_2\simeq 10^7W$ sec^{-1}. It is clear that the second level is populated by cascade and depopulated by ionization slowly enough for collisions to set up a fair approximation to the $2s-2p$ equilibrium. Among the higher n's, collisional redistribution is likely to be more efficient so that the statistical-weight distribution may be still more nearly approached,[3] but quantitative data are not yet available.

In any case recent work by SEATON and his collaborators suggests that the Balmer decrement is little affected by the resulting change in distribution among the sub-levels. Since we have chosen a case in which the $2p$ state behaves in effect quite similarly to the metastable $2s$ state, it is unlikely that the results of (iv) will require more than a very slight modification.

(vi) *The Importance of Collisional Excitation and De-excitation of Hydrogen in Emission Envelopes*

A further assumption, made in both (iii) and (iv), was that collisional excitation to the levels $n\geqslant 3$ is negligible in comparison with radiative processes. We shall now verify this assumption by means of the calculations of CHAMBERLAIN (1953), according to which collisional ionization and excitation of hydrogen tend to become important in gaseous nebulae when the electron temperature T_e appreciably exceeds 10,000°. We shall use his results to compute the relative frequencies of various collisional and radiative processes in shells, bearing in mind that the electron density N_e is about 10^9 times higher and the dilution factor W about 10^{12} times higher than in a typical planetary.

The equation of statistical equilibrium for the nth level in Case B may be written

$$\sum_{n'=2}^{n-1}F_{nn'}+\int_{\nu_n}^{\infty}F_{n\kappa}d\nu+\sum_{n'=n+1}^{\infty}F_{nn'}+\mathcal{F}_{n1}$$

$$=\sum_{n'=2}^{n-1}F_{n'n}+C_n+\sum_{n'=n+1}^{\infty}F_{n'n}+\mathcal{F}_{1n}. \qquad \ldots\ (36)$$

[3] M. J. SEATON, private communication.

where the F's represent radiative transitions, the $\mathcal{F}$'s represent collisional transitions, and terms of the form $\mathcal{F}_{nn'}$, $\mathcal{F}_{n'n}$ ($n' \neq 1$) are neglected.

Dealing first with the right-hand side of equation (36), and considering the recombination term which is independent of W, we have, neglecting the Gaunt factor,

$$C_n = 3{\cdot}26 \times 10^{-6} N_e^2 n^{-3} T_e^{-\frac{3}{2}} e^{\chi_n/kT_e} E_1(\chi_n/kT_e) \quad \ldots . (37)$$

(Cillié, 1932); while

$$\mathcal{F}_{1n} = N_1 N_e \int_{v_0}^{\infty} Q_n(v) v f(v) dv = N_1 N_e \alpha_n(T_e), \quad \ldots . (38)$$

where $\frac{1}{2} m v_n^2 = \chi_n$ and $\alpha_n(T_e)$ is tabulated by Chamberlain (1953). To compare (37) and (38), we need to find N_1, which we do by writing down the equation for statistical equilibrium of the ground level in the presence of diluted black-body radiation of colour temperature T_* and neglecting excitation by Lyman line radiation, 3-body recombinations and stimulated captures:

$$\int_{v_0}^{\infty} F_{1\kappa} dv + N_1 N_e \alpha_\kappa(T_e) = C_1,$$

or

$$N_1 \left[W \frac{8\pi}{c^2} \alpha \nu_0^3 E_1(h\nu_0/kT_*) + N_e \alpha_\kappa(T_e) \right] = 3{\cdot}26 \times 10^{-6} N_e^2 T_e^{-\frac{3}{2}} e^{h\nu_0/kT_e} E_1(h\nu_0/kT_e)$$

or

$$N_1 \left[6{\cdot}26 \times 10^9 W E_1\left(\frac{157820}{T_*}\right) + N_e \alpha_\kappa(T_e) \right] = 2{\cdot}07 \times 10^{-11} N_e^2 T_e^{-\frac{1}{2}} \quad \ldots . (39)$$

if $T_e \leqslant 40{,}000°$.

Assuming $T_* = T_e = T$ (which is a plausible approximation in the absence of forbidden lines), the two terms on the left of (39) may be deduced from Table VI.

Table VI

T	$\alpha_\kappa(T_e)$	$6{\cdot}26 \times 10^9 \, E_1\left(\frac{157820}{T}\right)$
10,000°	$2{\cdot}6 \times 10^{-14}$	15·0
20,000°	$5{\cdot}3 \times 10^{-12}$	$3{\cdot}0 \times 10^5$
40,000°	$6{\cdot}2 \times 10^{-10}$	$2{\cdot}4 \times 10^7$

Taking as typical values $N_e = 10^{12}$, $W = 10^{-2}$, the radiative ionization term exceeds the collisional term by over 100 if $T_* > 15{,}000°$, and this predominance will be greatly accentuated if, as the calculations of Baker, Menzel and Aller (1938) suggest, T_e is a few thousand degrees less than T_*. Hence collisional ionization from the ground level will be negligible except at the boundary of a Strömgren H II region where W must be replaced by $W e^{-\tau}$ with $\tau \sim 10$; this region, however, is too small to contribute appreciably to the emission.

In the interior of the ionized region, we thus obtain

$$\frac{N_1}{N_e^2} = (3{\cdot}31 \times 10^{-21}/T_e^{\frac{1}{2}}) W E_1\left(\frac{157820}{T_*}\right) = \\ = 4{\cdot}14 \times 10^{-16} e^{157820/T_*} T_*^{-1} T_e^{-\frac{1}{2}} W^{-1}, \quad \ldots . (40)$$

which is equivalent to PANNEKOEK's ionization equation.

We are now in a position to compare C_n and $\mathcal{F}_{1n}$. The ratio is

$$\frac{\mathcal{F}_{1n}}{C_n} = \frac{N_e N_1 \alpha_n(T_e)}{3{\cdot}26 \times 10^{-6} N_e^2 n^{-3} T_e^{-\frac{3}{2}} e^{\chi_n/kT_e} E_1(\chi_n/kT_e)} = \\ = 1{\cdot}27 \times 10^{-10} \frac{N_e}{W} n^3 \alpha_n(T_e) \frac{T_e}{T_*} \frac{e^{\chi_1/kT_*}}{e^{\chi_n/kT_e} E_1(\chi_n/kT_e)}. \quad \ldots . (41)$$

Assuming again $T_e = T_*$ (which naturally gives an upper limit to the relative importance of collisions), we have the following values of

$$\frac{W}{N_e} \frac{\mathcal{F}_{1n}}{C_n}:$$

Table VII

n \ T	10,000	20,000	40,000
2	85×10^{-16}	$6{\cdot}62 \times 10^{-16}$	$2{\cdot}03 \times 10^{-16}$
3	3·88	0·94	0·53
4	1·45	0·43	0·28
5	0·62	0·25	0·21
6	0·39	0·18	0·16

Below the broken line, $\mathcal{F}_{1n}/C_n < 1$ per cent if $W/N_e = 10^{-14}$. Collisional excitation actually increases in importance with decreasing temperature, owing to the rise in N_1, but naturally decreases very rapidly if T_e is reduced for a *given* T_*. Thus, if we interpolate in the results of BAKER, MENZEL and ALLER (1938) to find T_* corresponding to the above values of T_e for an optically thin nebula, we have the following for $(W/N_e)(\mathcal{F}_{1n}/C_n)$:

Table VIII

n \ T_* / T_e	11,000 / 10,000	22,000 / 20,000	49,000 / 40,000
2	18×10^{-16}	$2{\cdot}9 \times 10^{-16}$	$0{\cdot}80 \times 10^{-16}$
3	0·97	0·42	0·21
4	0·31	0·19	0·11
5	0·13	0·11	0·08
6	0·08	0·08	0·06

Even with these slight differences between T_* and T_e, the ratios $\mathcal{F}_{1n}/C_n$ are greatly reduced; the greatest value reached is 18 per cent for $n = 2$, $T_* = 11{,}000°$, $W/N_e = 10^{-14}$, and the other values are of order 1 per cent or less. Thus, even without considering the additional radiative processes entering equation (36), we may conclude that collisional excitation from the ground state is negligible in comparison with other processes populating each level.

Collisional excitation from excited levels.—The cross-sections for collisional excitation from, say, the second quantum level to higher levels are unknown. To estimate their order of magnitude very crudely, we may take the values of $\alpha_n(T_e)$ and augment them by the relevant BOLTZMANN factor $10^{51410/T_e}$. We thus have

$$\mathcal{F}_{2n} \sim 10^{51410/T_e} N_2 N_e \alpha_n(T_e) \sim \frac{b_2}{b_1} \mathcal{F}_{1n}.$$

In a shell that is optically thick in Lyman α but transparent to the Balmer and higher series, both b_1 and b_2 are proportional to $1/W$. One may therefore expect $\mathcal{F}_{2n}$ to be comparable with $\mathcal{F}_{1n}$ in order of magnitude. Terms $\mathcal{F}_{3n}$ etc., will be very much smaller, because b_3, etc., are less sensitive to W and generally somewhere in the neighbourhood of unity if the medium is transparent to radiation arising from the depopulation of these levels (see (iv)).

Collisional de-excitation—Detailed-balancing arguments give the following relation between the collisional de-excitation coefficient $\beta_n(T_e)$ to the ground level and the corresponding excitation coefficient $\alpha_n(T_e)$.

$$\beta_n = \frac{g_1}{g_n} e^{X_{1n}/kT_e} \alpha_n = \frac{1}{n^2} 10^{685380(1-1/n^2)/T_e} \alpha_n. \qquad \ldots . (42)$$

To ascertain the importance of collisional de-excitation for $n > 2$, it is convenient to compare the probability $N_e\beta_n$ of such a process, per atom in the nth level, with the spontaneous transition probability

$$\sum_{n'=2}^{n-1} A_{nn'}.$$

β_n is almost independent of T_e.

Table IX

n	$\sum_{n'=2}^{n-1} A_{nn'}$	β_n		
		$T_e = 10^4$	2×10^4	4×10^4
2	0	$8{\cdot}1 \times 10^{-9}$	$8{\cdot}3 \times 10^{-9}$	$11{\cdot}2 \times 10^{-9}$
3	$4{\cdot}4 \times 10^7$		$8{\cdot}3 \times 10^{-10}$	
4	$1{\cdot}7 \times 10^7$		$1{\cdot}9 \times 10^{-10}$	
5	$7{\cdot}4 \times 10^6$		$5{\cdot}4 \times 10^{-11}$	
6	$3{\cdot}5 \times 10^6$	$2{\cdot}6 \times 10^{-11}$	$2{\cdot}2 \times 10^{-11}$	$2{\cdot}8 \times 10^{-11}$

Table IX shows that, for $N_e = 10^{12}$, $N_e\beta_n$ is only of the order of $10^{-5}\Sigma A_{nn'}$, i.e. collisional de-excitation is completely negligible for $n \geqslant 3$.

For the second quantum level, there is, by hypothesis (Case B), no depopulation by spontaneous transitions. The next most convenient term on the left-hand side of equation (36) (and indeed the only one unless there is a significant radiation density in the Balmer lines) is that corresponding to photo-ionization, which is given by the following expression:.

$$\int_{\nu_0}^{\infty} F_{2\kappa} d\nu \simeq WN_2 \frac{8\pi\alpha\nu_0^3}{c^2} E_1(\chi_2/kT_*) \frac{1-(1/b_2)e^{-\chi_2/kT_e}}{1-e^{-\chi_2/kT_*}}$$

$$\simeq 2{\cdot}22\times 10^8 E_1\left(\frac{39440}{T_*}\right) \frac{1-(1/b_2)e^{-\chi_2/kT_e}}{1-e^{-\chi_2/kT_*}} . WN_2. \qquad \ldots . (43)$$

Assuming $T_* = T_e$, $b_2 = 1$ (both of which set lower limits to $F_{2\kappa}$), we have the following:

T	10^4	2×10^4	4×10^4
$\frac{\int F_{2\kappa} d\nu}{WN_2}$	$8{\cdot}6\times 10^5$	$1{\cdot}1\times 10^7$	$5{\cdot}0\times 10^7$

For $T = 10^4$, $N_e/W = 10^{14}$, collisional de-excitation of the second level is just comparable with photo-ionization at 10,000°, but fades to insignificance at the higher temperatures.

We conclude that, in general, one is justified in neglecting collisional excitation and de-excitation if $N_e/W \leqslant 10^{14}$ for T_e equal to or slightly less than T_*. P Cygni and AG Pegasi, for example, are both well within this limit. At higher densities or more extreme dilutions, or if T_* falls to 10,000° or less, collisional excitation and de-excitation processes may well take on importance, especially for the second level.

(vii) *Statistical Equilibrium of Helium in Envelopes and the Emission Curve of Growth: Be Stars*

Most of the absorption lines of helium in the spectra of Be stars are broad and diffuse, suggesting axial rotation with a visible component $v \sin i$ of the order of 200 km/sec. In some cases, however, such as the B3 spectroscopic binary ζ Tauri, He I λ 3965 ($2\ ^1S - 4\ ^1P$) is narrow and sharp, as are the absorption lines of Fe II, Ni II, Si II and the cores of the Balmer lines (Struve and Wurm, 1938). The broad lines, and the broad Balmer wings, presumably come from the underlying star, while the narrow ones probably arise in a shell at a distance of 2 or 3 stellar radii. He I λ 3965 is formed predominantly in the shell because its lower state $2\ ^1S$ is metastable and therefore heavily enhanced in population in the presence of diluted radiation; He I λ 3889 ($2\ ^3S - 3\ ^3P$) also arises from a metastable level and probably behaves similarly. A survey of several Be stars led Struve and Wurm to the following generalizations:

1. In the cooler stars, λ 3965 is the first absorption line to appear in the shell as the excitation is increased.

2. The triplet lines ($2\ ^3P - n\ ^3D$) appear in the shell when the number of atoms is large, but for the cooler stars λ 3965 is stronger than λ 4026 ($2\ ^3P - 5\ ^3D$).

3. When the excitation of the shell is higher (no Fe II absorption), the singlet lines ($2^1P - n^1D$) begin to appear faintly in the shell spectrum (e.g. $\lambda 4009$), but are generally fainter than $\lambda 3965$. In a normal B5 giant, on the other hand, $\lambda 4009$ is slightly stronger than $\lambda 3965$.

4. When the excitation in the shell is very high (as in P Cygni), the triplets become relatively stronger and $\lambda 4026$ surpasses $\lambda 3965$. The other singlets also become stronger in absorption (though remaining weaker than the triplets), and the metastability effect on $\lambda 3965$ is less noticeable. On the other hand, the singlets (other than $\lambda 3965$) are very weak in emission, compared with the triplets, so that this behaviour of the absorption lines must be due to curve-of-growth effects.

To account for these results, STRUVE and WURM considered the statistical equilibrium of an idealized helium atom having six states, $1\ ^1S$, $2\ ^1S$, $2\ ^1P$, $2\ ^3S$, $2\ ^3P$, and the ground state of He^+, for temperatures of 10^4 and $2{\cdot}5 \times 10^4$ respectively ($T_* = T_e$). Their calculations showed that the metastable levels $2\ ^1S$, $2\ ^3S$ both increase in relative population as W diminishes; this effect is more intense for $2\ ^3S$ because there is some leakage away from $2\ ^1S$ via $2\ ^1P$. $2\ ^1P$ has a strong transition to the ground state and diminishes very rapidly (roughly as W), while $2\ ^3P$ diminishes somewhat less rapidly. The effects of collisional excitation from the ground state were shown to be negligible for $N_e \leqslant 10^{16}$ cm^{-3}, even if $T_e = T_*$, though collisions between excited levels may perhaps play a part. The results were found to be in good agreement with observation for

$$W \simeq 0{\cdot}01; \quad R \simeq 5R_*.$$

Since some physical dilution is undoubtedly present, this value of R may be somewhat too large.

These statistical-equilibrium calculations were extended by GOLDBERG (1941) to an 8-level helium atom for the case $W = 0$ and by WELLMANN (1952a) to an atom with 22 levels for various values of W, T_*, and T_e; in each case the higher levels were introduced in an approximate way (b_i = const.). WELLMANN found the ratio of singlet to triplet populations to be less by a factor of 10 than that obtained by STRUVE and WURM.

In a further paper, WELLMANN (1952b, 1955) developed a curve-of-growth method for the application of these results to the observed intensities of emission lines, taking the effect of self-absorption into account. The integrated intensity of an emission line of mean frequency ν_0 emitted from a cylindrical column of unit area and length H is

$$I_{ji} = \int_0^\infty I_\nu d\nu = \frac{2h\nu_0^3}{c^2}\left(1 + \frac{W}{e^{h\nu_0/kT_*} - 1}\right)\frac{g_i}{g_j}\int_{\nu=0}^{\infty}\int_{S=0}^{H}\frac{N_j}{N_i}e^{-\tau_\nu}d\tau_\nu d\nu, \qquad \ldots . (44)$$

where j, i are the upper and lower levels with statistical weights g_j, g_i respectively Assuming N_j/N_i to be constant,

$$I_{ji} = \frac{2h\nu_0^3}{c^2}\left(1 + \frac{W}{e^{h\nu_0/kT_*} - 1}\right)\frac{b_j}{b_i}e^{-h\nu_0/kT_e}A_\nu, \qquad \ldots . (45)$$

where

$$A_\nu = \int_0^\infty (1 - e^{\varkappa_\nu N_i H}) d\nu. \qquad \ldots . (46)$$

The term in W is normally negligible and A_ν is simply the equivalent width of an absorption line formed in a tube with negligible re-emission. The relationship between

$$\frac{A_\nu}{\Delta\nu_D} \text{ and } \int \frac{N_i f_{ij} H}{\Delta\nu_D}, \qquad \text{or simply } \frac{A_\nu}{\nu} \text{ and } \frac{N_i f_{ij} H}{\nu}$$

is the standard curve of growth for exponential absorption (UNSÖLD, 1955). Neglecting differences of damping constants, this curve of growth should give a unique relation between

$$A' = \frac{E}{\nu^4} \frac{b_i}{b_j} e^{h\nu/kT_e} = \text{const.} \frac{A_\nu}{\nu} \qquad \ldots . (47)$$

and

$$N' = b_i g_i e^{\chi_i/kT_e} \frac{f_{ij}}{\nu} \frac{N_+}{g_+} = \text{const.} \frac{N_i f_{ij} H}{\nu} \qquad \ldots . (48)$$

from the horizontal shifts of which the relative abundances of different ions may be determined.

In practice, the effects of self absorption normally turn out to be small (WELLMANN, 1952c); when they are large, the assumption of constant N_j/N_i is naturally rather inaccurate. There is, however, a further effect tending to flatten or reverse the final line profile: the absorption of continuous stellar radiation in the portion of the envelope between the star and the observer. Inserting the Einstein transition coefficients A_{ji}, B_{ji}, B_{ij} and the line profile

$$\varphi(\nu) = \frac{\varkappa_\nu}{\int_0^\infty \varkappa_\nu d\nu}, \qquad \ldots . (49)$$

the flux, relative to the flux F_κ in the neighbouring continuum, becomes

$$\frac{F(\nu)}{F_\kappa} = 1 + \left[N_j \left(A_{ji} + B_{ji} \frac{4\pi}{c} W I_F \right) \eta - N_i B_{ij} \frac{4\pi}{c} W I_F \right] \frac{h\nu V}{4\pi^2 R_*^2 F_\kappa} \varphi(\nu), \qquad \ldots . (50)$$

where πI_F is the net flux of stellar radiation at frequency ν, η is a geometrical factor of the order of $(1 - W)$ to allow for occultation by the central star, and V the volume of the envelope. The net emission equivalent width is

$$w_\nu = \int_0^\infty \left[\frac{F(\nu)}{F_\kappa} - 1 \right] d\nu = \alpha' \frac{N_+ N_e V}{4\pi^2 R_*^2 F_\kappa} (b_j D_{ji} \eta - b_i D_{ij}) h\nu, \qquad \ldots . (51)$$

where

$$\alpha' = \frac{1}{2g_+}\frac{8\pi^2e^2}{mc^3}\left(\frac{h^2}{2\pi mkT_e}\right)^{\frac{3}{2}},$$

and

$$D_{ij} = g_i f_{ij} \nu_{ij}^2 e^{\chi_i/kT_e} W/(e^{h\nu/kT_*} - 1),$$

$$D_{ji} = g_i f_{ij} \nu_{ij}^2 e^{\chi_j/kT_e}\left(1 + \frac{W}{e^{h\nu/kT_*} - 1}\right).$$

D_{ji} and D_{ij} are proportional to the transition rates from j to i and from i to j respectively. Since D_{ij} is proportional to W, w_ν rapidly increases from negative values (absorption) to positive ones (emission) as W is reduced (Rosseland cycle), if b_i changes slowly. For metastable levels, on the other hand, b_i is nearly proportional to W^{-1}, so that the increase to net emission does not take place so rapidly. Thus, on the whole, the absorption tends to take place nearer to the star than the emission. The use of b_i's calculated on the assumption of purely geometrical dilution of the stellar line radiation may be justified in first approximation by reference to Rottenberg's scattering theory, provided that occultation can be neglected, if the internal motions are not too large.

The above method was applied by Wellmann (1952c) to the analysis of the well-known Be star γ Cassiopeiae, which showed a very well-developed He I spectrum in 1937–38. In the latter half of 1937, just after the star had apparently reached a maximum radius of 10 $R_\odot$ and a minimum colour temperature of 8,000°, the emission lines were single, narrow, and very intense. Shortly afterwards, at the end of 1938 January, the lines began to grow unsymmetrical, being fainter on the violet side. Towards the end of 1938, the emissions were split by central "shell" absorption lines, with the violet emission component fainter than the red one. In mid-1939, the emissions were all considerably weaker and the shell absorption lines much stronger, especially the hydrogen lines and the helium triplets; in metal lines, and helium singlets, the emission and absorption components either were both very weak or cancelled each other. These phenomena presumably represented the formation and subsequent disappearance of a large extended envelope.

Wellmann applied his curve-of-growth analysis to the emission lines of Fe II, H, and He I. Assuming b_i = const. for Fe II, the intensities fitted a linear relation with a tolerable scatter (due either to inaccurate f-values or to variations by factors of 2 or so in the b_i's) for electron temperatures of either 10,000° or 25,000°. A good linear fit was obtained for hydrogen, assuming either the b_n's of Case A_2 (Baker and Menzel, 1938) or those calculated with the inclusion of excitations from the ground state by absorption of stellar Lyman radiation (Case C: Baker, Menzel and Aller, 1938). A good fit was likewise obtained for He I, using the b_i's calculated for $W = 0{\cdot}01$ in agreement with the results obtained by Struve and Wurm for ζ Tauri, etc. Only λ 3888 ($2\,^3S - 3\,^3P$) departs seriously from the curve, owing to the effect of the absorption of stellar radiation from the metastable $2\,^3S$ level, according to (51). The intensities of the other lines with metastable lower states suggest $T_e \simeq 10{,}000°$.

By combining the curves for H and He I through a lateral shift, Wellmann showed that the Balmer decrement must actually correspond to Case A_2 and not

Case C, in view of the normal relative abundances of hydrogen and helium. There is considerable evidence that Be stars do not differ from ordinary B stars in chemical composition (BURBIDGE and BURBIDGE, 1951; STRUVE, 1951; BOYARTCHUK, 1958). Case B also seems to be excluded by the same argument (unless hydrogen is underabundant in γ Cas) and by the complete absence of self-absorption effects even in Hα. We have seen in (iv) that, for P Cygni and AG Pegasi, Case B appears to be the most likely one; however, our conclusion that there is no excitation by (Balmer) line radiation from the star in these cases is entirely analogous to WELLMANN's conclusion that there is no excitation by (Lyman) line radiation in γ Cas. In each case this result is understandable in terms of the presence of strong absorption lines (either "stellar" or due to the innermost part of the shell) at the relevant wavelengths in the spectrum of the star. The exciting frequencies for He I, on the other hand, are present in nearly the amounts expected from the geometrical dilution factor, although, of course, there will be a limited degree of physical dilution. Enough helium is present in the metastable levels to produce appreciable self-absorption in the strongest He I lines $\lambda 5875$ ($2\,^3S - 3\,^3D$) and $\lambda 7065$ ($2\,^3P - 3\,^3S$), although the high rotational velocity ($v \sin i = 420$ km/sec at the stellar surface) ensures effective transparency in all other cases.

WELLMANN proceeds to discuss the case in which the density and dilution factor vary with distance from the stellar surface. The statistical-equilibrium calculations (WELLMANN, 1952a) show that, approximately, $b_i \propto 1/W$ for the metastable levels and $b_i \simeq$ const. for most of the others. (51) may now be rewritten

$$L = \frac{4\pi^2 R_*^2 F_\kappa \nu_\nu}{h\nu} = 4\pi\alpha' \int_{R_*}^{\infty} (b_j D_{ji}\eta - b_i D_{ij}) N_e N_+ R^2 dR, \qquad \ldots . (52)$$

where

$$\eta' = (1 - W).$$

For $\lambda 5015$ ($2\,^1S - 3\,^1P$), for example, (52) becomes

$$L_{5015} = 4\pi\alpha' \left[\overline{b_j D_{ji}\eta'} - \bar{c}_i \overline{\left(\frac{D_{ij}}{W}\right)} \right] \int_{R_*}^{\infty} N_e N_+ R^2 dR, \qquad \ldots . (53)$$

where

$$c_i = W b_i$$

and the bar indicates a mean value over slowly varying quantities. The first term in brackets in (53) is virtually equal to $\overline{b_{ij}} D_{ji}$, since D_{ji} and η' are insensitive to W when the latter is small.

Assuming an exponential density law

$$N_e N_+ = \text{const. } e^{-\mu R/R_*}, \qquad \ldots . (54)$$

the $\overline{b_j}$'s found from various He I line ratios were substituted into the definitions of the mean

$$x_1 = \frac{\int b(3^3P) N_e N_+ R^2 dR}{\int N_e N_+ R^2 dR}$$
$$x_2 = \frac{\int W b(2^3P) N_e N_+ R^2 dR}{\int N_e N_+ R^2 dR} \quad \ldots . (55)$$

etc., to obtain μ from the ratios of $\lambda\lambda$ 7065, 6678, and 5875 to λ 5015. The two values are quite consistent, with a mean $\mu = 3{\cdot}4$.

The corresponding values of W range from 0·05 to 0·14, somewhat higher than obtained in the initial analysis according to (47), (48) in which the absorption terms in (51) were neglected. Their inclusion leads to an improved curve of growth shifted to the left by about 0·5 along the log N' axis. A comparison of the new curve of growth for He I with that for hydrogen (for which the absorption effects are small) gives

$$\log \frac{N_{\mathrm{He}^+}}{N_{\mathrm{H}^+}} = -0{\cdot}90. \quad \ldots . (56)$$

Thus He may have a degree of ionization anywhere between $\frac{1}{2}$ and unity, according to the abundance assumed.

Wellmann now finds N_e by assuming an absolute magnitude $M_{\mathrm{vis}} = -3^{\mathrm{m}}{\cdot}2$ and colour index $\mathrm{CI}_{\mathrm{int}} = -0^{\mathrm{m}}{\cdot}2$, from which

$$4\pi^2 R_*^2 F_\kappa = 10^{33{\cdot}0} \text{ ergs/sec/A at } \lambda 4000. \quad \ldots . (57)$$

From the equivalent width of Hα, assuming $\overline{b_3} = 0{\cdot}1$ and $R_* = 15 R_\odot$,

$$N_e N_{\mathrm{H}^+} = 10^{23{\cdot}6} e^{-3{\cdot}4 R/R_*} \quad \ldots . (58)$$

or

$$N_e = 10^{11{\cdot}8} e^{-1{\cdot}7 R/R_*}, \quad \ldots . (59)$$

assuming $q = 1$. N_e can also be estimated from the continuous optical depth of the shell; if this arises from electron scattering the "veiling" of the stellar absorption lines during the period of strong emission requires

$$\tau = 0{\cdot}2 = \sigma_e \int_{R_*}^{\infty} N_e dR,$$

whence

$$N_e = 10^{12{\cdot}4} e^{-1{\cdot}7 R/R_*}. \quad \ldots . (60)$$

The agreement between (59) and (60) is satisfactory in view of the uncertainties involved. The corresponding optical depth for the Lyman continuum is quite uncertain owing to the uncertainty in T_* and T_e. Wellmann concludes that the envelope must derive its support predominantly from rotation, the effects of turbulence and radiation pressure being negligible in comparison with gravity. It follows that the envelope must be more like a ring than a spherical shell.

Further results were obtained by WELLMANN from the spectrum of the declining phase of the shell towards the end of 1938, when the absorption lines were more prominent. In particular, the emission lines of hydrogen continued to lie on the linear portion of the curve of growth calculated according to (51) even when the absorptions were strong, and the curve of growth for the latter implied that the "shell" absorption lines are definitely formed from a stellar continuum unmodified by emission lines. In other words, the absorption lines are formed in the deepest layer of the shell and are then partially filled in by overlying emission. A comparison of Hβ and He I λ 4471 suggested an electron density 500 times greater than in the maximum phase and $W \simeq 1$. The increased density was thus roughly compensated by a reduction in the extent of the envelope. Still later (1939 August), the excitation temperature began to rise steeply as the envelope presumably fell back into the "star proper".

A somewhat different picture of the relative stratification of emission and absorption in the envelope was obtained by BOYARTCHUK (1957) from a study of γ Cas and 10 other Be stars observed in the Crimea between 1953 and 1956. Assuming constant angular momentum throughout the envelope, the rotational broadening of the shell absorption and emission lines (compared with that of the underlying stellar absorption) suggested $R \simeq 1{\cdot}7\, R_*$ ($W = 0{\cdot}1$) for the emission lines and $R \simeq 3{\cdot}5\, R_*$ ($W = 0{\cdot}02$) for the absorptions. This result must depend on the line examined: as has already been mentioned, the population of metastable levels (including the $2s$ level of hydrogen when collisional $2s - 2p$ transitions are rare) increases as $W \to 0$. Thus for the Balmer series and for several of the He I lines this result is quite natural; it would be most interesting to study the rotational broadening of lines arising from lower levels that are not metastable. From the curve of growth of the hydrogen absorption lines, BOYARTCHUK found optical depths similar to those estimated for P Cygni in Table V.

4. The Transfer of Radiation in Stellar Envelopes

In the investigations described in Section 3, it was assumed that the populations of the various levels—as expressed by b_n or b_i—were known and either constant (Cases A, B for hydrogen) or simple functions of W (Case C for He I); they were assumed to be uninfluenced by the transfer of radiation either because the effective optical depth is small or because scattered radiation is re-emitted at a sufficiently nearby frequency to be available for further excitations. We now discuss the effect of absorption and re-emission on the populations of the various excited levels.

(i) *Transfer in a Stationary Envelope*

The first accurate treatment of the transfer of continuous radiation through a grey spherical shell, or of line radiation through a stationary spherical shell with coherent scattering, was given by MILNE (1930). If the shell is thin compared with its radius and far enough away from the central star for the direct radiation from the latter to be considered as concentrated in a parallel beam, MILNE showed that the spherical problem could, with the appropriate boundary conditions, be reduced to a plane-parallel problem. If τ is the radial optical depth from the exterior surface of the shell to a point within the shell, the attenuation of the initial flux πF incident on

the inner boundary is exactly compensated by the fluxes associated with the diffuse outward and inward beams $I(\tau, \theta)$, $I(\tau, \psi)$ respectively, where

$$0 \leqslant \begin{matrix}\theta \\ \psi\end{matrix} \leqslant \pi/2.$$

Thus the combined equation of transfer and radiative equilibrium may be written

$$\int_0^1 I(\tau, \mu_\theta)\mu_\theta d\mu_\theta - \int_0^1 I'(\tau, \mu_\psi)\mu_\psi d\mu_\psi + \tfrac{1}{2}Fe^{-(\tau_1-\tau)} = \tfrac{1}{2}F, \qquad \ldots . (1)$$

where τ_1 is the total radial optical depth of the shell and $\mu_\theta = \cos\theta$, $\mu_\psi = \cos\psi$. The corresponding boundary conditions are

$$I'(0, \mu) = 0; \; I(\tau_1, \mu) = I'(\tau_1, \mu). \qquad \ldots . (2)$$

Combined with the equations of transfer for I and I', these lead in Eddington's approximation to the expression for the source function

$$B(\tau) = \tfrac{1}{2}F(1 + \tau + 2e^{-\tau_1} - e^{-(\tau_1-\tau)}), \qquad \ldots . (3)$$

which corresponds to that in a stellar atmosphere when $\tau_1 \to \infty$. This model predicts limb-brightening for small values of τ_1, which may in principle be measured either by observing this phenomenon or by comparing the brightnesses of the stellar and nebular images. The ratio of the total brightnesses is simply

$$\frac{\text{Light from star}}{\text{Light from nebula}} = \frac{e^{-\tau_1}}{1 - e^{-\tau_1}}. \qquad \ldots . (4)$$

Applying (4) for example to Kron and Gordon's luminous shell surrounding the WR component of V444 Cygni, we have $\tau_1 = 0{\cdot}7$ if, as appears to be the case, the core and disk have about the same total luminosity. This value of τ_1 leads to an appreciable limb-brightening (the maximum intensity, at $\mu = 0{\cdot}3$, exceeds the central intensity by 30 per cent), in agreement with Kron and Gordon's conclusion that the residuals are reduced by postulating a bright ring with a radius of the order of $2R_*$.

The next step forward was made by Ambarzumian (1933), who discussed the transfer of Lyman continuum radiation in a geometrically thin stationary hydrogen nebula in which the primary fluorescence mechanism operates, eventually converting each quantum beyond the Lyman limit into a Lyman-α quantum and a quantum of a higher series (which immediately escapes). If p is the probability that an absorbed ultra-violet quantum is re-emitted by a recombination to the ground state ($p \simeq \frac{1}{2}$), the equation of radiative equilibrium in the Schuster-Schwarzschild approximation is

$$p[I(\tau) + I'(\tau) + \tfrac{1}{2}Fe^{-(\tau_1-\tau)}] = 2B(\tau). \qquad \ldots . (5)$$

Together with the transfer equation, this leads to the result

$$\pi F_u = 0{\cdot}7\pi F e^{-\tau_1} \qquad \ldots . (6)$$

for the emergent flux of ultra-violet diffuse radiation if $\tau_1 \geqslant 3$. The corresponding flux of Lyman-α *photons* is

$$\frac{\pi F}{h\nu_c}(1-1{\cdot}7e^{-\tau_1}) \simeq \frac{\pi F}{h\nu_c} \qquad \ldots\ldots (7)$$

and the final energy flux emerging in Lyman α is simply

$$\pi F_\alpha = \frac{\nu_\alpha}{\nu_c}\pi F,$$

where ν_α, ν_c are the frequency of Lyman-α and the mean frequency in the Lyman continuum respectively. By considering the transfer of Lyman-α through the nebula, Ambarzumian derived the following expression for the mean intensity in Lyman-α at the inner boundary (optical depth t_1):

$$\tfrac{1}{2}[I(t_1)+I'(t_1)] = \frac{\nu_\alpha}{\nu_c}\frac{t_1}{\tau_1}(\tau_1-1)F. \qquad \ldots\ldots (8)$$

Since $t_1/\tau_1 \simeq 10^4$, the mean intensity for $\tau_1 = 2$ is about $10^4 F$ or 4×10^4 times the mean intensity of Lyman continuum radiation incident from the star. The resulting excitation to the second quantum level is likely to lead to an appreciable optical depth in the Balmer series. Furthermore, since the absorption coefficient for Lyman-α is much higher than for Lyman continuum radiation, the flux $(\nu_\alpha/\nu_c)\pi F$ incident on the outer parts of the nebula leads to a radiation-pressure gradient about 10^4 times that of gravity and tends to cause expansion of the outer parts relative to the inner parts.

In shells around early-type stars, however, the situation is somewhat different, since the geometrical dilution factor is not low enough for the Rosseland cycle to be operative (Miyamoto, 1950). However, the presence of motion in the atmosphere probably requires some re-examination of this conclusion, since excitation to the higher quantum levels will be reduced in consequence.

When the nebula or envelope has internal motions comparable with or larger than the Doppler broadening parameter of the line absorption coefficient, the effective optical depth is greatly reduced by the "leakage" effect. This is so even in the case of uniform expansion, since the component of the expansion velocity along the direction of propagation of radiation will in general be variable.

From his calculation of the radiation pressure of Lyman-α, Ambarzumian (1933b) showed that a stationary finite nebula cannot be in equilibrium. He therefore considered radiative transfer in a moving shell transparent to radiation incident from other layers, replacing the boundary condition (2) by

$$I'(0) = I(t_1) = 0. \qquad \ldots\ldots (9)$$

Radiation from the central star in the neighbourhood of the line frequency is thus neglected. The algebraic difference between the Lyman-α fluxes at the inner and outer boundaries of the nebula is still $(\nu_\alpha/\nu_c)\pi F$, although their actual values depend on the continuous optical thickness τ_1. Hence there is still a strong tendency for the nebular material to be dissipated under the conditions assumed. A more elaborate and precise treatment of the same problem was given by Chandrasekhar (1935), but Zanstra (1949) showed that the radiation pressure of Lyman-α is reduced by a

factor of 300 when the Gaussian line profile and non-coherent scattering are taken into account.

The transfer of Lyman radiation in a hydrogen envelope surrounding an emission-line star, in which the dilution factor cannot be considered to be infinitely small, was considered in a further paper by AMBARZUMIAN (1935). Whereas the intensity of Lyman-α given by (8) is of the order of 10^{-4} times that at the stellar surface (apart from line absorption in the stellar atmosphere, which of course has no effect on (8)), so that the number of transitions F_{23} is negligible in comparison with F_{13}, the corresponding energy density in a stellar envelope is very much higher, so that F_{23} must be taken into account. This will be so if $\tau_1 > 1$; for $\tau_1 < 1$, there is little conversion of Lyman continuum into Lyman-α, so that the theory of Section 3 remains applicable. On the other hand, even if τ_1 is large, cyclic processes cannot be neglected in comparison with ordinary scattering, since their presence is indicated by the presence of the emission lines. AMBARZUMIAN considered an idealized atom with two discrete energy levels and a continuum and solved two linked transfer equations for the line and continuous radiations. He obtained the result:

$$\frac{N_e N_+}{N_1} = \frac{g_+}{g_1}(2\tau+1)W\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} e^{-h\nu_{13}/kT} \qquad \ldots . (10)$$

provided that $10^{-6} < W < 10^{-2}$. In a stationary shell, the ionization is thus enhanced in comparison with the usual value assumed, neglecting ionizations from the second level, by the factor $(2\tau+1)e^{(\tau_1-\tau)}$, which may be very considerable. However, this formula is of doubtful applicability in a moving envelope, since the population of the excited level will be reduced by the leakage effect.

The variation of the electron temperature through an envelope that is optically thick in the Lyman continuum is discussed by ALLER, BAKER and MENZEL (1939).

(ii) *Transfer in a Moving Envelope*

The importance of motion in the envelope for the transfer of radiation and excitation of atoms in all cases where the line optical depth is appreciable has already been indicated in (i). A detailed treatment of this problem for stellar envelopes was given by V. V. SOBOLEV (1947) in his remarkable treatise on "Moving Envelopes of Stars",[4] the first part of which will now be discussed. This presents a theory of a homogeneous medium in which the equations of transfer are reduced to simultaneous non-differential equations comparable to the statistical-equilibrium equations introduced for nebulae by CILLIÉ and by MENZEL and his co-workers, and applies it to the emission-line spectra of early-type stars. The medium is assumed to be transparent in the Lyman continuum, and to have a uniform isotropic velocity gradient. SOBOLEV shows that this situation is considerably simpler than that of a stationary envelope, since radiation is absorbed and re-emitted only in a limited region in which the velocity variation is comparable with the Doppler broadening parameter of the line, so that no reference need be made to conditions at the boundaries of the envelope. The medium can thus be considered as an infinite one, with a constant radiation density in the ionizing continuum.

Owing to the velocity gradient in the medium, a certain fraction β_{ki} of the radiation emitted from a downward transition $i \to k$ escapes, while the rest is absorbed and

[4] A translation of SOBOLEV's treatise, by S. GAPOSCHKIN, will be published shortly by the Harvard Press.

effects the reverse transition in a nearby atom. For example, in BAKER and MENZEL's Case A, $\beta_{12} = 1$ and, in Case B, $\beta_{12} = 0$. Hence the equation of statistical equilibrium for level i is simply

$$N_i\left(\sum_{k=1}^{i-1} A_{ik}\beta_{ki} + B_{i\kappa}\rho_{i\kappa}\right) = \sum_{k=i+1}^{\infty} N_k A_{ki}\beta_{ik} + N_e N_+ C_i(T_e), \quad \ldots . (11)$$

where $\rho_{i\kappa}$ is the density of ionizing radiation from the ith level given in terms of the density of the same radiation at the stellar surface by

$$\rho_{i\kappa} = W\rho_{i\kappa}^* \quad \ldots . (12)$$

and $B_{i\kappa}$ is the corresponding EINSTEIN transition coefficient. To determine β_{ki}, we use an argument given by ZANSTRA (1956). Suppose the absorption coefficient is given by the Doppler formula

$$\alpha_\nu = \alpha e^{-(w/w_D)^2} = \alpha e^{-x^2}, \quad \ldots . (13)$$

where α is the absorption coefficient at the line centre, with frequency ν_0 as measured in some fixed frame of reference. Suppose the systematic velocity in the line of sight is

$$v = \frac{dv}{ds}s = aw_D s, \quad \ldots . (14)$$

where s is the optical depth in the line centre from a point where $v = 0$, and a is the (constant) ratio of the velocity-gradient to the product of α and the Doppler parameter. If we observe at a fixed frequency $\nu_0 + \Delta\nu$, the corresponding optical depth may be found by integrating along the line of sight over all elements for which

$$c\frac{\Delta\nu}{\nu} = v + w$$

and, if the envelope is "infinite", this condition may equally well be replaced by

$$w = -v = -aw_D s. \quad \ldots . (15)$$

Hence the optical depth is

$$\tau(\Delta\nu) = \int_{-\infty}^{\infty} e^{-a^2 s^2} ds = \frac{\sqrt{\pi}}{a}. \quad \ldots . (16)$$

For non-coherent scattering with a constant source-function J, the total intensity produced in optical depth τ is τJ, while the emergent intensity is $J(1 - e^{-\tau})$. Hence

$$\beta_{ki} = \frac{1}{\tau}(1 - e^{-\tau}) = \frac{a}{\sqrt{\pi}}(1 - e^{-\sqrt{\pi}/a}) \quad \ldots . (17)$$

$$\simeq a/\sqrt{\pi} \quad \ldots . (18)$$

if $a < 1$. The corresponding expression obtained by SOBOLEV, assuming a rectangular

line profile, is $a/2$, which is not significantly different from the approximate expression (18). When self-absorption is small, however, $a > 1$ and the more accurate expression (17) must be used.

The simple expression (18) has the advantage that $A_{ik}\beta_{ki}$ can be expressed in terms of $A_{21}\beta_{12}$, using the standard relations between A_{ik} and α_{ki}. The relation is

$$A_{ki}\beta_{ik} = \frac{\frac{g_2}{g_1}N_1 - N_2}{\frac{g_k}{g_i}N_i - N_k}\left(\frac{\nu_{ik}}{\nu_{12}}\right)^3 A_{21}\beta_{12}. \qquad \ldots . (19)$$

Using (19), the equation (11) becomes

$$N_i\left[x\sum_{k=1}^{i-1}\frac{\frac{g_2}{g_1}N_1 - N_2}{\frac{g_i}{g_k}N_k - N_i}\left(\frac{\nu_{ki}}{\nu_{12}}\right)^3 + \frac{B_{i_K}}{A_{21}}\rho_{i_K}{}^*\right] \qquad \ldots . (20)$$

$$= x\sum_{k=i+1}^{\infty}N_k\frac{\frac{g_2}{g_1}N_1 - N_2}{\frac{g_k}{g_i}N_i - N_k}\left(\frac{\nu_{ik}}{\nu_{12}}\right)^3 + \frac{N_e N_+ C_i(T_e)}{WA_{21}},$$

where

$$x = \beta_{12}/W = \frac{1}{w_D\sqrt{\pi}\alpha_{12}}\frac{dv}{dR}\frac{1}{W}. \qquad \ldots . (21)$$

Equation (20) takes account of stimulated emissions in the lines, but not of ROTTENBERG's effect of excitation by stellar radiation at the line frequency. Apart from this factor, however, it accurately describes the statistical equilibrium of all levels for which $a_{ik} < 1$ for the most significant members of the series. For higher levels, $\beta_{ik} \to 1$ and the "nebular" statistical-equilibrium equations must be used. An attractive feature of (20) is that the relative populations N_i/N_k depend only on the ratio of the velocity gradient and dilution factor, and not on these quantities separately.

SOBOLEV solves equation (20) for a hydrogen atom with 6 levels and continuum, assuming self-absorption in all series, for the cases $T_e = T_* = 20{,}000°$ (P Cygni, etc.) and $T_e = T_* = 50{,}000°$ (WR stars), and for $x = 0$, 0·01, 0·1, 1·0 and 10. The same results apply to He^+ at temperatures four times as high as these. The ionization $N_e N_+/N_1 W$ turns out to be close to the equilibrium value for all x. The populations of the various levels have $b_i = 1/W$ for $x = 0$ (complete absorption), but diminish rapidly as x increases; for example N_2/N_1 diminishes by a factor of 30 as x increases from 0 to 1. SOBOLEV's values for N_3/N_1, etc., are somewhat inaccurate, (a) because of the limited number of levels assumed, and (b) because the stimulated emission factor has apparently been neglected in the transitions to the continuum; this is important for $i > 3$.

SOBOLEV proceeds to calculate the Balmer decrement ($H\beta = 1$) on the basis of these results, using the formula

$$\frac{E_{k2}}{E_{42}} = \frac{N_k A_{k2}\beta_{2k} h\nu_{2k}}{N_4 A_{42}\beta_{24} h\nu_{24}} = \frac{4\dfrac{N_2}{N_4} - 1}{\dfrac{k^2}{4}\dfrac{N_2}{N_k} - 1}\left(\frac{\nu_{2k}}{\nu_{24}}\right)^4, \qquad \ldots . (22)$$

and also deduces that the energy in the Balmer continuum is four times that in Hα for $T = 20{,}000°$, $x = 0{\cdot}1$. This result, and the Balmer decrement calculated from SOBOLEV's theory for absorption in all series with $x = 0{\cdot}1$, are so close to the corresponding results for the "nebular" Case B that it is impossible to distinguish between them observationally. SOBOLEV's claim that $x = 0{\cdot}1$ fits P Cygni better than the "nebular" decrement is based on a comparison of his calculation with that of CILLIÉ (1932), both of which are somewhat inaccurate owing to the limited number of levels taken into account.

We may, however, use the observations of the Paschen lines P12–17 in P Cygni carried out by BURBIDGE and BURBIDGE (1955) to distinguish between SOBOLEV's theory and the "nebular" case. According to SOBOLEV's theory, we have from (19) (neglecting stimulated emission)

$$\frac{E_{12,3}}{E_{12,2}} = \frac{A_{12,3}\beta_{3,12}\nu_{3,12}}{A_{12,2}\beta_{2,12}\nu_{2,12}} = \frac{g_3}{g_2}\frac{N_2}{N_3}\left(\frac{\nu_{3,12}}{\nu_{2,12}}\right)^4 = 0{\cdot}565, \qquad \ldots . (23)$$

while on the "nebular" theory with no self-absorption

$$\frac{E_{12,3}}{E_{12,2}} = \frac{A_{12,3}}{A_{12,2}}\frac{\nu_{3,12}}{\nu_{2,12}} = 0{\cdot}306. \qquad \ldots . (24)$$

As was shown in Section 3(vi), the ratio found from the equivalent widths is close to 0·31. To obtain a ratio of 0·56, one would have to assume a colour temperature for the underlying continuum of 9500°, which is unreasonably low in view of the large amount of interstellar reddening present (ROMAN, 1951). In this case, therefore, observation seems to favour the "nebular" hypothesis rather than that of SOBOLEV.

For WR stars, on the other hand, SOBOLEV obtains a much flatter decrement than the nebular one. Although the intensities of the hydrogen lines are hard to estimate in practice, owing to blending with the PICKERING series of He^+, he makes out a good case for assuming his decrement for hydrogen, together with the nebular decrement for He^+, in the spectrum of HD 192163 (BEALS, 1934). Again, the observed decrement in Ne and Se stars is much flatter than the nebular one and agrees with $x = 0{\cdot}01$.

When x is large, the populations of the higher levels become low, leading to transparency in the higher series. Accordingly, SOBOLEV gives an approximate method of estimating the level at which $a \to 1$. Neglecting stimulated emission, the equations (20) are added term by term, giving

$$\sum_{i=j}^{\infty} N_i\left(x\sum_{k=1}^{j-1} D_{ki}\frac{N_i}{N_k} + B_{i\kappa}\rho_{i\kappa}^*\right) = \frac{N_e N_+}{W} S_j, \qquad \ldots . (25)$$

where

$$D_{ki} = 4A_{21}\frac{g_k}{g_i}\left(\frac{\nu_{ki}}{\nu_{12}}\right)^3, \quad S_j = \sum_{i=j}^{\infty} C_i. \qquad \ldots . (26)$$

Since N_i and D_{ki} decrease rapidly with increasing i, (25) may be replaced by the recurrence formula

$$N_j\left(x\sum_{k=1}^{j-1} D_{kj}\frac{N_1}{N_k} + B_{j_K}\rho_{j_K}{}^*\right) = \frac{N_e N_+}{W} S_j. \qquad \ldots . (27)$$

From (27), N_1 etc. can be determined to find at which level $\beta_{ik} \geqslant 1$.

In particular, it follows from (14) and (26) that, if

$$\frac{A_{43}}{D_{34}}\frac{N_3}{N_1} < \beta_{12} < \frac{A_{32}}{D_{23}}\frac{N_2}{N_1}, \qquad \ldots . (28)$$

i.e., if

$$2 \times 10^{-4} < \beta_{12} < 3 \times 10^{-3}$$

for $x = 1{\cdot}0$, then the medium is opaque in the Balmer series but transparent in the Paschen and higher series. Sobolev solves the equations (25) for this case and obtains for $x \sim 10$ a very high ratio of Hα to Hβ, owing to the degradation of the higher Balmer lines by fluorescence into Hα and members of higher series. Such an Hα : Hβ ratio has been observed in the post-maximum spectra of novae and is likely to be due to this cause, as is the occurrence of emission in Hα alone in stars such as α Cygni. The values of x required for these various cases are quite consistent with the dimensions and expansion velocities of the various envelopes deduced from observation.

Sobolev proceeds to apply his theory to a discussion of physical conditions in the envelopes surrounding early-type stars with bright lines. He first obtains the condition at the lower boundary of the envelope, at which the photospheric level corresponds to an optical depth due to electron scattering of 1/3 in the radial direction. Together with the inverse square law for the electron density, this gives

$$N_e{}^{(0)}R_* = 0{\cdot}5 \times 10^{24}, \qquad \ldots . (29)$$

which may be compared with equation (3) of Section 1. The neglect of neutral hydrogen absorption is justified for supergiants with $T_* \geqslant 20{,}000°$ because the optical depth beyond the Lyman limit is small enough to permit a high degree of ionization. A further relation between N_e and R is obtained from the intensity of any given Balmer line:

$$E_{k2} = A_{k2}h\nu_{2k}\int N_k\beta_{2k}dV = \frac{16}{k^2}\left(\frac{\nu_{2k}}{\nu_{12}}\right)^3 A_{21}h\nu_{2k}\int_{R_*}^{\infty} f_{k2}N_e N_+ 4\pi R^2 dR, \ldots . (30)$$

where (neglecting induced emissions)

$$f_{k2} = \frac{N_k}{N_2}\frac{WN_1}{N_e N_+}x. \qquad \ldots . (31)$$

f_{k2} turns out to be substantially independent of x and equal to about $1{\cdot}5 \times 10^{-21}$ if $T = 20{,}000°$. Hence, from the measured equivalent widths,

$$qN_e^{(0)^2}R_* = \frac{k^2 H_{k2}\pi}{8c^2 A_{21} f_{k2}}\frac{\nu_{12}^3}{e^{h\nu_{2k}/kT}-1}, \qquad \ldots . (32)$$

where H_{k2} is the line intensity w_ν/ν as defined by ZANSTRA. Equation (32) may be compared with our equation (14) in Section 3(iii). By combining (29) and (32), SOBOLEV obtains $R_* = 20R_\odot$, $N_e^{(0)} = 3{\cdot}6 \times 10^{11}$ for a typical Be or P Cygni star and $R_* = 5R_\odot$, $N_e^{(0)} = 1{\cdot}5 \times 10^{12}$ for a WR star. These results are in fairly good agreement with those given earlier, if it is assumed that AG Pegasi may be classed (for this purpose) as a WR star. From these figures, it follows that WR and P Cygni stars both lose mass at a rate of $10M_\odot$ in 10^6 years, while Be stars lose mass at about $0{\cdot}1$ of this rate. Actually it is questionable whether Be stars are losing mass at all and whether the inverse square law of density variation applies to them (BOYARTCHUK, 1957).

From its definition, it follows that x increases as $(R/R_*)^3$, so that the excitation decreases very rapidly as one goes out into the envelope (constant expansion velocity). SOBOLEV finds the radial optical depth in the centre of Hα to reach the value $1/3$ at a radius $R_1 = 12R_*$ in the case of P Cygni stars and $4R_*$ in WR stars, so that the "shell" absorption lines are formed very far out, but not as far out as would be inferred from our Table V. It also follows from SOBOLEV's theory that the optically thin region makes only a small contribution to the intensities of the emission lines.

This last point leads to a very elegant explanation of the *profiles* of the lines in P Cygni and Be stars. It has already been mentioned in Section 3(i) that the emission-line profile of a uniformly expanding spherical envelope—obtained by calculating the volumes of elements having the same velocity component in the line of sight—is simply rectangular if random motions are neglected. Such a profile appears to be observed in certain nova shells and WR stars, but is by no means common. To obtain the observed profiles of P Cygni emission lines on the "optically thin" hypothesis, it is necessary to make special assumptions about the velocity gradients and random motions. If, on the other hand, a considerable degree of self-absorption is present, the intensity at a given point of the profile is proportional to the *surface area* (projected on the plane normal to the line of sight) of a cone with its apex at the star and semi-angle θ, where

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}\cos\theta. \qquad \ldots . (33)$$

If random motions are neglected, each cone-surface may be regarded as transparent to the radiation from another cone-surface owing to the leakage effect. Hence the intensity at a given point of the profile is simply

$$I_\lambda \propto \sin^2\theta \propto 1 - \frac{c^2}{v^2}\left(\frac{\Delta\lambda}{\lambda}\right)^2, \qquad \ldots . (34)$$

giving a parabolic profile. Such a profile gives an excellent representation of the P Cygni emission component in BEALS' Type I.

Similarly, the depression at a given point of the absorption component is proportional to the amount of light from the star obscured by the cone-surface whose velocity in the line of sight corresponds to that frequency. Thus the residual intensity is

$$r(\Delta\lambda) = 1 - \frac{\int_{v}^{\theta_1} \psi(\theta) \cos\theta \sin\theta \, d\theta}{\int_{0}^{\pi/2} \psi(\theta) \cos\theta \sin\theta \, d\theta} \qquad \ldots . (35)$$

where $\psi(\theta)$ is the limb-darkening function and

$$\begin{aligned} \sin\theta_1 &= \frac{R_1}{R_*} \sin\theta \qquad && \text{if } R_1 \sin\theta < R_* \\ \theta_1 &= \pi/2 \qquad && \text{if } R_1 \sin\theta \geqslant R_*, \end{aligned} \qquad \ldots . (36)$$

R_1 being the radius of the opaque region. Numerical computations based on equation (30) and reasonable guesses at $\psi(\theta)$ give the characteristic asymmetrical line contour with its centre displaced to the violet; this calculation is essentially similar to that of ROTTENBERG (Section 2(iv)), but, in contrast to ROTTENBERG's theory, the emission and absorption lines are here simply superposed. It is characteristic of SOBOLEV's theory that the residual intensity in the absorption line is due to the unobscured portion of the star rather than to re-emission in the envelope; hence both the limb-darkening law and the extent of the opaque region are of fundamental importance in fixing the line profile.

For the Be profiles, SOBOLEV expresses in quantitative form MCLAUGHLIN's hypothesis of combined rotation, pulsation, and absorption. The surfaces of equal radial velocity now assume a somewhat more complex form, but the direct superposition of the emission and absorption components in this case readily reproduces MCLAUGHLIN's red-violet asymmetry.

SOBOLEV's theory thus gives an elegant explanation of the observed line profiles in Be and P Cygni stars. Furthermore, there can be little doubt that self-absorption is often present in the lowest members of the Balmer series. On the other hand, observation suggests that there is little self-absorption in the higher members of the Balmer series in stars such as P Cygni, so that it is not at all clear that SOBOLEV's theory will always give better results for the populations of the various levels than the straightforward application of BEALS' hypothesis, in spite of its obvious successes in accounting for line profiles and for certain "anomalous" Balmer decrements.

With these reservations, it appears that the theories of SOBOLEV and of ROTTENBERG offer the most promising means of explaining many features of the emission spectra of early-type stars. Although their predictions are in certain cases remarkably similar, the two theories are in a sense complementary to one another, since ROTTENBERG's is based on excitation by line radiation from the star (which SOBOLEV neglects) and ignores fluorescence effects (except as a basic source of line radiation that is

then coherently scattered). There is also the mathematical difference that ROTTENBERG concentrates on the boundary conditions at the shell and neglects transfer effects, while SOBOLEV does the exact opposite; this difference, however, is probably less important than the physical difference mentioned above, owing to the inevitable uncertainty as to the true density distribution in an envelope. The two theories predict similar line profiles in the simpler cases largely because the emission and absorption are treated as surface, rather than volume, effects; but SOBOLEV's theory is really more complete because it makes definite predictions of the population in each level as a function of the parameter x which is roughly calculable, while ROTTENBERG has to assume a given amount of fluorescent radiation to start with in order to normalize his predicted profiles in accordance with the observed intensities. It is, however, clear that ROTTENBERG's mechanism is essential to account for the more complex P Cygni profiles noted by BEALS. Thus it appears that a complete analysis of an emission spectrum must in general make use of all three hypotheses of BEALS, SOBOLEV and ROTTENBERG in order to account for the intensities and profiles of the lines. For this purpose, it would be of great utility to solve SOBOLEV's equations of statistical equilibrium, extended to allow for the presence of absorption in lower members of a series with transparency in higher members of the same series. This task will be a simple one for an electronic computer.

REFERENCES

ALLEN, C. W.	1955	*Astrophysical Quantities* (Athlone Press, London) p. 174.
ALLER, L. H.	1943	*Ap. J.*, **97**, 135.
	1954	*Astrophysics II*, (Ronald Press Co., New York), p. 157.
	1956a	*Gaseous Nebulae* (Chapman and Hall, London), p. 135.
	1956b	*ibid.*, p. 261.
ALLER, L. H., BAKER, J. G. and MENZEL, D. H.	1939	*Ap. J.*, **90**, 601.
AMBARZUMIAN, V. A.	1933a	*M.N.*, **93**, 50.
	1933b	*Bull de l'Obs. Central à Poulkova*, **13**, no. 3
	1935	*M.N.*, **95**, 469.
BABCOCK, H. W.	1958	*Ap. J. Supp.*, **3**, 200.
BAKER, J. G. and MENZEL, D. H.	1938	*Ap. J.*, **88**, 52.
BAKER, J. G., MENZEL, D. H. and ALLER, L. H.	1938	*Ap. J.*, **88**, 422
BEALS, C. S	1930	*Publ. Dominion Astrophys. Obs., Victoria*, **4**, 294.
	1934	*Publ. Dominion Astrophys. Obs., Victoria*, **6**, 95.
	1940	*J. Roy. Astron. Soc. Canada*, **34**, 169.
	1944	*M.N.*, **104**, 205.
	1950	*Publ. Dominion Astrophys. Obs., Victoria*, **9**, no. 1.
BEALS, C. S. and HATCHER, R. D.	1948	*Contrib. Dominion Obs. Ottawa*, **2**, no. 1.
BEALS, C. S. and ROTTENBERG, J. A.	1956	*Vistas in Astronomy*, II, ed. A. Beer (Pergamon Press, London), p. 1437.
BOWEN, I. S.	1928	*Ap. J.*, **67**, 1.
	1934	*Publ. Astron. Soc. Pacific*, **46**, 146.
	1935	*Ap. J.*, **81**, 1.
BOYARTCHUK, A. A.	1957	*Astronomichesky Zhurnal*, **34**, 193.
	1958	*Mem. Soc. Roy. des Sciences de Liège*, **20**, 159.
BURBIDGE, E. M. and BURBIDGE, G. R.	1951	*Ap. J.*, **113**, 84.
	1955	*Ap. J.*, **122**, 89.
BURBIDGE, G. R. and BURBIDGE, E. M.	1952	*Ap. J.*, **116**, 185.
	1954	*Ap. J.*, **120**, 76.
CHAMBERLAIN, J. W.	1953	*Ap. J.*, **117**, 387.
CHANDRASEKHAR, S.	1934a	*M.N.*, **94**, 444.
	1934b	*M.N.*, **94**, 522
	1935	*Z. Astrophys.*, **9**, 266.
	1947	*Ap. J.*, **106**, 145.

Cillié, G.	1932	*M.N.*, **92**, 820.
Eddington, A. S.	1926	*Proc. Roy. Soc., A.*, **111**, 429.
Gaposchkin, S.	1941	*Ap. J.*, **93**, 202.
Gerasimovič, B. P.	1933	*Z. Astrophys.*, **7**, 335.
Giovanelli, R. G.	1948	*Austral. J. sci. Res. A.*, **1**, 289.
Goldberg, L.	1941	*Ap. J.*, **93**, 244.
Hiltner, W. A.	1945	*Ap. J.*, **101**, 356.
	1950	*Ap. J.*, **112**, 477.
Jager, C. de	1956	*Solar Eclipses and the Ionosphere*, ed. W. G. Beynon and G. M. Brown (Pergamon Press, London), p. 174.
Keeping, E. S.	1947	*Publ. Dominion Astrophys. Obs., Victoria*, **7**, 349.
Kogure, T.	1953	*Pub. Astron. Soc. Japan*, **5**, 1.
Kopal, Z.	1944	*Ap. J.*, **100**, 204.
Kopal, Z. and Shapley, M. S.	1946	*Ap. J.*, **104**, 160.
Kosirev, N. A.	1934	*M.N.*, **94**, 430.
Kourganoff, V.	1958	*Mém. Soc. Roy. des Sci. de Liège*, **20**, 117.
Kron, G. E. and Gordon, K.	1943	*Ap. J.*, **97**, 311.
	1950	*Ap. J.*, **111**, 454.
McLaughlin, D. B.	1931	*Publ. Amer. Astron. Soc.*, **7**, 31.
	1933	*Proc. Amer. Acad. Sci.*, **19**, 44.
	1956	*Vistas in Astronomy*, II, ed. A. Beer (Pergamon Press, London), p. 1477.
Menzel, D. H.	1937	*Ap. J.*, **85**, 330.
Menzel, D. H., Aller, L. H., Baker, J. G., Goldberg, L., Hebb, M. H. and Shortley, G. H.	1937/45	*Physical Processes in Gaseous Nebulae* (reprinted from *Ap. J.*).
Menzel, D. H., Aller, L. H. and Baker, J. G.	1938	*Ap. J.*, **88**, 313.
Merrill, P. W.	1951a	*Ap. J.*, **113**, 605.
	1951b	*Ap. J.*, **114**, 338.
Milne, E. A.	1928	*M.N.*, **89**, 3.
	1930	*Z. Astrophys.*, **1**, 98.
Miyamoto, S.	1950	*Ap. J.*, **113**, 181 and *Contrib. Inst. Astrophys. Kyoto, passim.*
	1952	*Publ. Astron. Soc. Japan*, **4**, 37.
	1953	*Publ. Astron. Soc. Japan*, **5**, 55.
Münch, G.	1950	*Ap. J.*, **112**, 266.
Neubauer, F. J. and Aller, L. H.	1948	*Ap. J.*, **107**, 281.
Oke, J. B.	1954	*Ap. J.*, **120**, 22.
Pagel, B. E. J.	1958	*Mém. Soc. Roy. des Sci. de Liège*, **20**, 177.
Pannekoek, A.	1926	*Bull. Astron. Inst. Netherlands*, **3**, 207.
Pelt, A. van	1957	*Bull. Astron. Inst. Netherlands*, **13**, 285.
Roman, N. G.	1951	*Ap. J.*, **114**, 492.
Rosseland, S.	1926	*Ap. J.*, **63**, 218.
Rottenberg, J. A.	1952	*M.N.*, **112**, 125.
Schuster, A.	1905	*Ap. J.*, **21**, 1.
Seaton, M. J.	1955	*Proc. Phys. Soc.*, A, **68**, 457.
Sherman, F. and Morgan, W. W.	1939	*Ap. J.*, **89**, 509.
Sobolev, V. V.	1947	*Moving Envelopes of Stars*, Publ. Leningrad State Univ. (in Russian). English Transl. by S. Gaposchkin; in press (at the Harvard University Press).
Strömgren, B.	1935	*Z. Astrophys.*, **10**, 237.
	1939	*Ap. J.*, **89**, 526.
Struve, O.	1942	*Ap. J.*, **95**, 134.
	1944	*Ap. J.*, **100**, 189.
Struve, O.	1951	*Astrophysics*, ed. J. A. Hynek (at the University Press, Chicago), p. 85.
Struve, O. and Roach, F. E.	1939	*Ap. J.*, **89**, 509.
Struve, O. and Wurm, K.	1938	*Ap. J.*, **88**, 84.
Swings, P.	1942	*Ap. J.*, **95**, 112.
Swings, P. and Struve, O.	1940a	*Publ. Astron. Soc. Pacific*, **52**, 392.
	1940b	*Ap. J.*, **91**, 546.
	1942	*Ap. J.*, **96**, 254.
Thackeray, A. D.	1955	*M.N.*, **115**, 236.
	1956	*Vistas in Astronomy* II, ed. A. Beer (Pergamon Press, London), p. 1380.
Tcheng Mao-Lin and Bloch, M.	1952	*Ann. Astrophys.*, **15**, 104.
Thomas, R. N.	1947	*Ap. J.*, **106**, 482.
	1949	*Ap. J.*, **109**, 500.

UNDERHILL, A. B.	1948	*Ap. J.*, **107**, 247.
	1949	*Ap. J.*, **110**, 340.
	1950	*Kgl. Danske Vid. Selsk. Mat. Fys. Medd.*, **25**, No. 13.
	1958a	*Mém. Soc. Roy. des Sci. de Liège*, **20**, 21.
	1958b	*ibid*, p. 23.
UNSÖLD, A.	1955	*Physik der Sternatmosphären* (Berlin: Julius Springer), p. 288.
WEENEN, J.	1950	*Bull. Astron. Inst. Netherlands*, **11**, 176.
WELLMANN, P.	1952a	*Z. Astrophys.*, **30**, 71.
	1952b	*Z. Astrophys.*, **30**, 88.
	1952c	*Z. Astrophys.*, **30**, 96.
	1955	*Vistas in Astronomy* I, ed. A. Beer (Pergamon Press, London), p. 303.
WILSON, O. C.	1934	*Ap. J.*, **80**, 259.
	1940	*Ap. J.*, **91**, 379.
	1942	*Ap. J.*, **95**, 402.
WOOLLEY, R. v. D. R.	1934	*M.N.*, **94**, 361.
ZANSTRA, H.	1927	*Ap. J.*, **65**, 50.
	1931a	*Publ. Dominion Astrophys. Obs. Victoria*, **4**, 209.
	1931b	*Z. Astrophys.*, **2**, 336.
	1932	*M.N.*, **93**, 131.
	1949	*Bull. Astron. Inst. Netherlands*, **11**, 1.
	1956	*Amsterdam Obs. Circ.*, No. 10.
ZANSTRA, H. and WEENEN, J.	1950	*Bull. Astron. Inst. Netherlands*, **11**, 165.

Colour, Luminosity and Evolution of the Stars

Olin J. Eggen

Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, Sussex

The Problem

"Fortunately, the data needed for the construction of these (stellar) histories are easily found. Our problem is like that of one who enters a forest of oaks, and desires to learn through what stages the trees have passed in reaching their present condition. He cannot wait long enough to see any single tree go through its long cycle of change. But on the ground he may find acorns, some unbroken and some sprouting. Others have given rise to rapidly growing shoots, and saplings are at hand to show the next stage of growth. From saplings to trees is an easy step. Then may be found, in the form of dead limbs and branches, the first evidence of decay, reaching its full in fallen trunks, where the hard wood is wasting to powder."

George Ellery Hale
The Study of Stellar Evolution,
1909.

Some Solutions

"And the Sun is by no means so dense as it may become . . . and will still continue in its (contraction), even if it only attained the density of the Earth—though it will probably become far denser in the interior owing to the enormous pressure—this would develop fresh quantities of heat, which would be sufficient to maintain for an additional 17,000,000 of years the same intensity of sunshine as that which is now the source of all terrestrial life."

Hermann von Helmholtz
Königsberg Lectures, 1854.

"I pointed out in the year 1886 that the time had arrived when stars with increasing temperatures would require to be fundamentally distinguished from those with decreasing temperatures, but I did not then know that this was so easy to accomplish as it now appears to be . . . I here introduce the 'temperature curve'. . . . On one arm of this we have those stages in the various heavenly bodies in which in each case the temperature is increasing, while on the other arm we have that other condition in which we get first vaporous combinations, and then ultimately the formation of a crust due to the gradual cooling of the mass in dark bodies like, say, the companion of Sirius."

Sir Joseph Norman Lockyer
Proceedings of the Royal Society,
Vol. 44, 1888.

"Again, no one supposes that the stars of Class B were *created* hot. They have doubtless originated in some way or other from more widely diffused primordial matter. Such primordial matter, if any still exists, must be relatively cold, for it is not luminous. Between it and the stage of spectrum B there must be intermediate stages in which the mass is growing hotter, while its

light is yellow or red, and, on account of its greater size, is much brighter than in the later stage of its history when it reaches the same temperature while cooling off."

HENRY NORRIS RUSSELL
The Observatory,
Vol. 37, 1914.

"It may, therefore, reasonably be maintained that the Draper (spectral) classification is not only convenient for descriptive purposes, but that the order in which the spectra have been placed represents the probable order of celestial evolution, such evolution taking place continuously from the chaotic conditions of nebulae, through the white, yellow, and red stars until luminosity ceases."

ALFRED FOWLER
The Observatory,
Vol. 38, 1915.

"It now appears that a typical star will be born as a super-giant, probably with a surface-temperature too low for visibility, and possibly as a variable undergoing violent pulsations. It will gradually move westward in our diagram until it reaches the region of high ionization which we have called the Pacific slope; when it reaches this region, it turns to the south-east and travels down the coast. At this period the centre of the star is a pool of fully ionized matter which is entirely inert as regards generation of energy. The extent of this pool will in time reach a maximum, after which it declines and finally becomes insignificant. When this stage is reached, the star leaves the Pacific coast and travels across the continent in a south-easterly direction, ultimately falling into the Atlantic."

SIR JAMES HOPWOOD JEANS
Monthly Notices,
Vol. 85, 1925.

"For the past 50 years stellar astronomy has been guided by successive theories of evolution. Today we have no theory of evolution pending the settlement of the laws of subatomic energy. . . . On the old theory of evolution from type B to M, and on the (RUSSELL) giant and dwarf theory of evolution from giant M up to B and down to dwarf M, the star's *track of evolution* was fixed by familiar physical laws. The laws of subatomic energy did not guide the star on the magnitude-type diagram, they settled its *rate of progress*. But now there is nothing left but the conditions of liberation of subatomic energy to guide the star; these must determine its track if it is evolving, or its halting place if it is stationary."

SIR ARTHUR STANLEY EDDINGTON
The Internal Constitution of The Stars, 1926.

THE MORAL

"We must be prepared to meet with side-tracks in evolution. Nature does not run in a groove. Her operations are free and various; they defy the restrictions of feasibility which a rigid methodism of thought would seek to impose. . . . Our best attitude of mind, then, in attempting to speculate on the course of things, is that of the utmost flexibility to the teaching of well-ascertained facts."

AGNES CLERKE

HALE's silvan analogy for stellar evolution, quoted above, is a perversion of a comparison made many years earlier by WILLIAM HERSCHEL. It is a perversion because HALE's version contraverts Sir William's main point, that is, without an assumed pre-knowledge of trees, we are unable to tell whether oaks become acorns saplings, shoots, brown leaves on the ground, green leaves on the tree; or, in each case, *vice versa*.

But abandoning dendrology, in what follows we will emphasize one approach to—or manifestation of—stellar evolution: the colour-luminosity array. From the observed magnitudes and colours of the (assumed) coeval members of star clusters or groups, we will attempt to give a résumé of the recent rapid progress in this approach. The observations discussed have been published by several observers—amongst the largest contributors are A. SANDAGE, R. H. STOY, N. G. ROMAN, G. E. KRON, M. WALKER and especially H. L. JOHNSON and his cohorts. On the theoretical side, most of the ideas here expressed (and here outraged into simplicity) have been proposed, altered or applied by A. SANDAGE, M. SCHWARZSCHILD, F. HOYLE, S. CHANDRASEKHAR, and others; an incomplete, but representative, bibliography is appended. The *avant-garde* of the last few generations of theoretical astrophysicists have been to a large extent concerned with stellar evolution and, indeed, together with the question of the cosmical distance scale, it might be considered the *raison d'être* of modern astronomy.

"The nebula called Praesepe", said GALILEO in 1610 in one of the first applications of the telescope to sidereal astronomy, "is not one star only, but a mass of more than forty small stars". Modern investigations of proper motions, complete to about

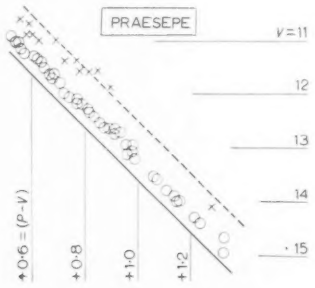


FIG. 1

visual magnitude 13·5 in a 2° square area, have picked out approximately 200 stars as members of this cluster. The magnitudes, V, and colours, P–V, of all these stars with P–V greater than $+0^m5$ are plotted in Fig. 1. Thirteen stars, which appear to stand above the run of the rest, are indicated by crosses; the remainder define a main sequence which can be represented by the linear relation,

$$V = 4^m84\,(P–V)+8^m40 \qquad \ldots\ (1)$$

If the observed magnitude and colour were due to a close double star, consisting of

two unresolved components of equal colour and magnitude, the observed colour would be that of the individual stars but the combined magnitude would place the observed point 0^m75 above the run of the single stars in Fig. 1, that is, on the broken line. Since the crosses fall either on this line, or between it and the main sequence, it is reasonable to suppose that in each case the observed magnitude and colour is that of a multiple system. If this is so, the apparent lack of a gradual transition from the main sequence to the broken line, and the near absence of double stars fainter than $V = 12^m$, merits further investigation.

Our next concern is to find the zero-point of Equation (1), that is, to convert the observed visual magnitudes to absolute values. There are seven dwarf stars within 5 parsecs of the Sun that have observed colours—undistorted by known companions—in the range covered in Fig. 1. The absolute visual magnitudes,

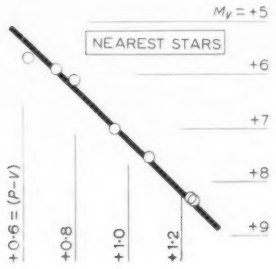


FIG. 2

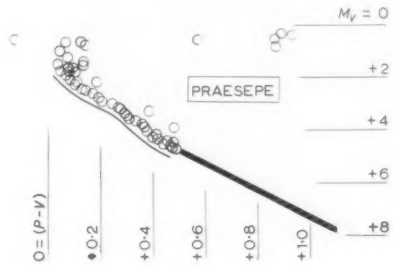


FIG. 3

M_V, of these stars are plotted in Fig. 2. When Equation (1), whose slope is represented by the shaded bar in the figure, is adjusted to fit these points, we find:

$$M_V = 2^m36 + 4^m84\ (\text{P–V}) \qquad \ldots\ (2)$$

or $V - M_V = +6^m04$. The distance of the Praesepe cluster is then 161 parsecs, or its parallax $\pi = 0''.0062$. This derivation of the distance rests on the fundamental assumption that both the slope and zero-point of the main sequences of the stars in Figs. 1 and 2 are identical; we shall return to this point.

An additional check on Equation (2) is afforded by the colour and luminosity of the Sun. From a recent investigation by STEBBINS and KRON we know that P–V $= +0^m53$ and, therefore, from Equation (2) $M_V = +4^m92$, whereas the direct determination by STEBBINS and KRON is $M_V = +4^m84$.

Figure 1 contains about one-half the members of the Praesepe cluster. Using the value of $V - M_V = +6^m04$ we can now plot the colour-luminosity array for the bluer members; this is shown in Fig. 3. Only stars for which photoelectric determinations of the magnitudes and colours are available have been plotted in Fig. 3—the dozen or so known cluster members excluded by this fact would add nothing to the array.

The colours, P–V, represent the ratio of the radiations received from two, slightly overlapping, segments of the stellar energy curves between 3800 and 6000 Å. A valuable, additional observational result, available for the Praesepe stars is the U–B colour—the ratio of the observed radiations from very slightly overlapping

segments of the energy curves on the violet (U) and red (B) sides of 3800 Å. The relationship between the two colours for the Praesepe stars is shown in Fig. 4 where it is clear that the Balmer discontinuity in the U-radiation from the stars bluer than approximately P–V = +0^{m}5 introduces a down-turn from the near linear relation valid for redder stars. The crosses in the figure, near P–V = +0^{m}9, represent the red giants in the upper right of Fig. 3.

It cannot be too strongly emphasized that the arrangement of the points in a colour-luminosity array such as Fig. 3 is strictly controlled by the colour system used, and before the colour co-ordinate can be interpreted in terms of, say, temperature, the complicated dependence upon such things as line-strength and variations in continuous absorption coefficient must be considered.

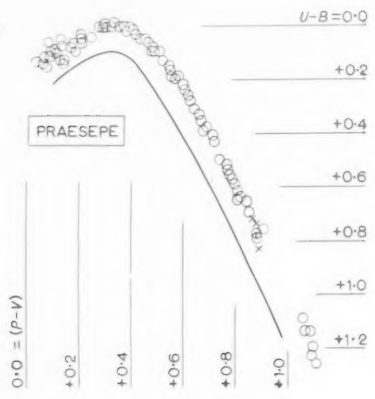


Fig. 4

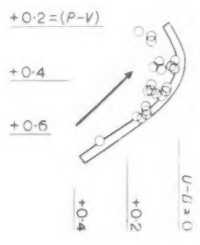


Fig. 5

Another cluster of interest in the present connection is the Pleiades. Although more distant than Praesepe and showing less proper motion, 190 Pleiades members brighter than photographic magnitude 16 and within 1° of the bright star Alcyone are known from an extensive investigation by Hertzsprung. Published photoelectric results, however, are only available for about one-half this number, nearly all of which are bluer than P–V = +0^{m}5. The stars for which P–V is greater than +0^{m}2 and for which (U–B)-colours are available are plotted in Fig. 5. The enclosed segment in the figure represents a part of the relationship between the two colours for the Praesepe stars in Fig. 4; since that part of this relationship between P–V = 0 and +0^{m}2 is not as well defined as the rest, probably because of the greater scatter in the luminosities of the stars represented (Fig. 3), it is not used in Fig. 5. The arrow in the figure indicates the slope of the reddening correction as determined by several investigations of highly reddened, mostly early-type stars, that is, the ratio (U–B)/(P–V) = 0·85 for the reddening effect of interstellar absorbing material on the two colours. There is some evidence that this ratio may not be universal, but the observed variations are small enough to be of no consequence for the amount of reddening considered here.

If we assume that the radiation actually emitted by the Pleiades members represented in Fig. 5 does not, in the wavelengths concerned, differ from that of the Praesepe stars used to define the closed segment in the figure, then the displacement

of the Pleiades stars from the standard relationship can be interpreted as due to interstellar absorption. From the greater scatter of the points in Fig. 5 than was found in Fig. 4 for the Praesepe stars in the same colour range, we might also surmise that the reddening is not uniform, but the differential effect must be small and the adoption of a mean reddening for the stars will introduce an amount of uncertainty unimportant to our present purpose. Mean reddenings of $E_{U-B} = +0^{m}06$ and $E_{P-V} = +0^{m}05$ have been subtracted from all the observed colours of Pleiades members and the results plotted in Fig. 6.

After the observed magnitudes were corrected for interstellar absorption, $V_0 = V-3E_{P-V}$, the stars with corrected values of P–V between $+0^{m}4$ and $+0^{m}5$ were fitted to the Praesepe main sequence in Fig. 3, giving a distance modulus of

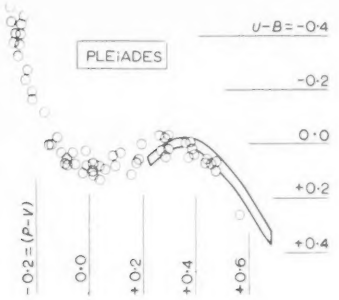


FIG. 6

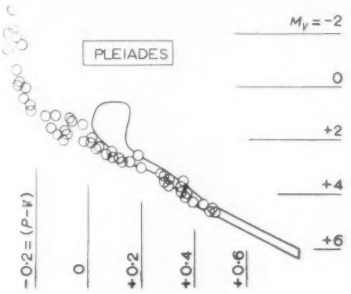


FIG. 7

$V_0-M_V = +5^{m}70$. The resulting colour-luminosity array for the Pleiades is shown in Fig. 7 where the enclosed area represents the array for Praesepe stars in Fig. 3. With the exception of a few stars above the main sequence near P–V = 0 to $-0^{m}1$, the Pleiades show a well defined main sequence extending to $M_V = -3^{m}$; two known visual binaries have been omitted from the figure.

The colour-luminosity arrays for three additional clusters, chosen for their contrast with the Pleiades and Praesepe, are shown in Figs. 9, 11 and 13. The relationships between the two colours P–V and U–B for these clusters are shown in Figs. 8, 10 and 12, where the standard, the Pleiades–Praesepe relationship is also outlined. In all three cases there is little reason to suspect that any appreciable reddening is affecting the observed colours. The main sequences of the Coma Berenices and M 67 clusters, in Figs. 9 and 11, were fitted to that of the Praesepe cluster in Fig. 3 for values of P–V greater than $+0^{m}4$; in NGC 752 the faintest stars between P–V = $+0^{m}3$ and $+0^{m}4$ were used to make this fit. Although many additional stars in M 67 have been observed photographically with results comparable in accuracy to the photoelectric observations, only the latter have been used in Figs. 12 and 13. The resulting distance moduli are $V-M_V = 4^{m}80$, $7^{m}78$, and $9^{m}50$, for Coma Berenices, NGC 752 and M 67, respectively.

Except for some clusters that have bluer, and brighter, main sequence stars than those of the Pleiades, the 5 galactic clusters whose colour luminosity arrays are shown in Figs. 3, 7, 9, 11 and 13 are representative examples of the known varieties.

Also, these examples offer the advantage of having been subjected to extensive astrometric investigations which permit us to separate the cluster members from the fore- and background stars. Indeed, the contribution of astrometry to the investigations summarized here cannot be over emphasized, and to a large degree the extension of these investigations now waits upon further astrometric work. The demands upon the astrometry become, of course, more severe because the nearer clusters,

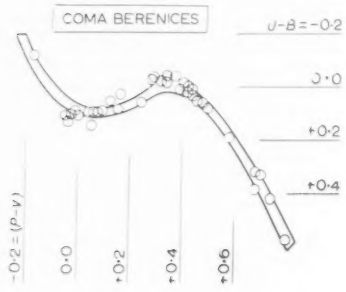


FIG. 8

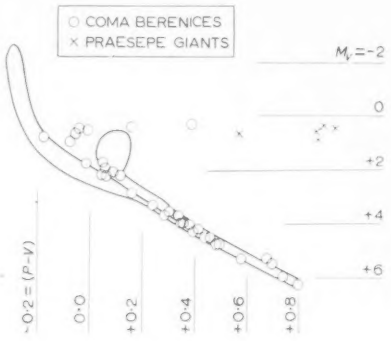


FIG. 9

which might be expected to show the greater proper motions, have been investigated first, leaving only the more distant ones with correspondingly smaller proper motions. It is true, however, that even the clusters discussed above have by no means been worked out from an astrometric standpoint, especially in the case of the fainter and outlying cluster members.

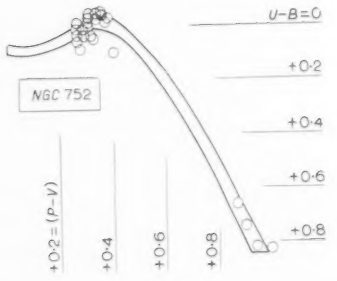


FIG. 10

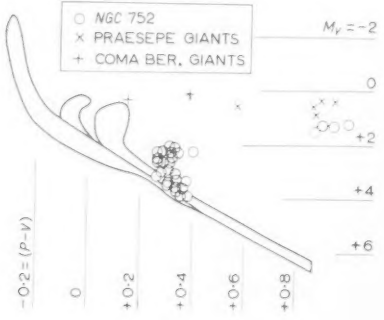


FIG. 11

The successive colour luminosity arrays in Figs. 5, 7, 9, 11 and 13 show, in outline, the section of the array populated by previously discussed clusters until, in Fig. 13, we see a regular progression of "break-away" points from the main sequence, ranging from P–V near -0^m2 for the Pleiades, through -0^m05 for Coma Berenices, $+0^m1$ for Praesepe, and $+0^m3$ for NGC 752 to $+0^m5$ for M 67. This striking feature and many others were first noted, in outline at least, by ROBERT TRUMPLER; using spectral types (H–R diagrams), TRUMPLER, working single-handed in the twenties, accomplished much of the work now being done, by many, with colour-luminosity arrays.

Another noteworthy fact, evident in the colour-luminosity arrays for Coma Berenices, Praesepe and NGC 752, is that the few giants which appear in each case show very little dispersion about a mean luminosity equal to or only slightly higher than that of the brightest early-type stars in the cluster. Although M 67 fits in, generally, with the run of the other clusters, the colour-luminosity array in Fig. 13 shows two

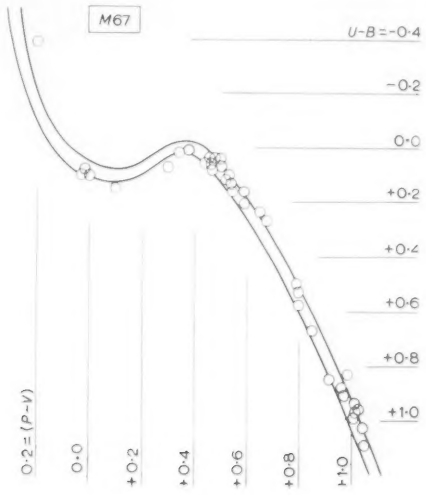


FIG. 12

distinguishing features; (a) the presence of several stars that are bluer than the break-away point and (b) a comparatively large dispersion in the luminosity of the giants. The colour-luminosity arrays for Praesepe and Coma Berenices in Figs. 3 and 9, however, also show in each case one star that is significantly bluer than the rest and feature (a) may result from the fact that, whereas the members of M 67, the

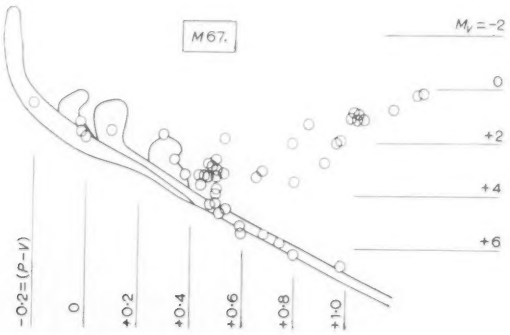


FIG. 13

richest galactic cluster known, are numbered by thousands, Coma Berenices and Praesepe contain approximately only 50 and 200 stars, respectively. The comparatively large number of giants in M 67 can also be explained by the richness of the cluster and, perhaps, this fact can also be invoked to explain the larger spread in luminosity of these giants since, when *all* the main sequence stars are considered, the

brightest giants observed in M 67 are still only slightly more luminous than the brightest early-type stars.

Attempts to interpret the colour-luminosity arrays discussed above have, usually, taken two directions in which both the similarities and the differences between one array and another are explained as the effect of,

A. Original chemical composition of the stars
B. Evolutionary history of the stars.

It is safe to predict that if the true answer contains features of either explanation, it contains features of both.

The origins of the modern views of stellar evolution are easily traced to the work of RUSSELL, and especially of EDDINGTON, a quarter of a century ago. A convenient and important milestone is the interpretation, mainly in terms of A above, of the HERTZSPRUNG-RUSSELL diagram published by BENGT STRÖMGREN in the *Zeitschrift für Astrophysik* of 1933: at this stage stellar evolution was introduced only as a correction to the main argument.

"We hereby arrive", said STRÖMGREN, "at the conclusion that mass and hydrogen content determine a star's configuration and appearance, and hence determine the star's position in the HERTZSPRUNG-RUSSELL diagram. . . . The peculiarities of the distribution (of stars) in the H–R diagram are a consequence of the peculiarities in the distribution of mass and hydrogen content and of the peculiarities of the transformation from the mass–hydrogen-content diagram to the H–R diagram." By

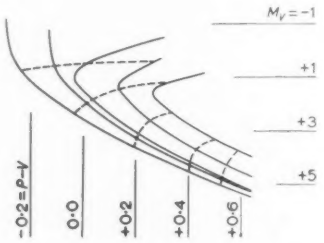


FIG. 14

making this latter transformation, from semi-empirical mass–hydrogen-content and purely empirical mass-luminosity diagrams, based on the material then available, STRÖMGREN constructed the lines of equal hydrogen content and equal mass in the H–R diagram. In Fig. 14 we have converted this result to the colour-luminosity array, where the continuous lines connect points of equal hydrogen content, the broken lines points of equal mass. A quantitative value for the hydrogen content, connected with each continuous curve, is not given and would have little meaning since (*a*) some of the basic assumptions made by STRÖMGREN have been altered by more recent work and (*b*) the exact form of the curves in Fig. 13 depend upon uncertain bolometric corrections and complicated temperature-colour relations.

The important result is the similarity of the continuous curves in Fig. 14 to the colour-luminosity arrays of the clusters discussed above, especially in the progression

of the break-away points from the main sequence as we pass from the Pleiades to M 67. Since this latter feature of the colour-luminosity arrays had already been noted by TRUMPLER in the cluster H–R diagrams, the possibility of explaining these cluster diagrams on the basis of STRÖMGREN's results was immediately investigated by G. P. KUIPER. In the *Astrophysical Journal* for 1937 KUIPER published his conclusions: "The hypothesis is proposed that *the clusters are groups of stars of nearly the same hydrogen content*. This hypothesis explains well-known properties of clusters. (Ordinary giants do not occur in a cluster that contains O or early B stars. . . . The spectrum-luminosity arrays of clusters show less scatter than the arrays for stars in general. . . . The main sequences for clusters differ in shape.) The composite cluster diagram is identified with a set of lines of constant hydrogen content, although systematic deviations may occur for the very luminous stars, owing to the ageing of the clusters. The relative position of the clusters is adjusted with . . . spectroscopic parallaxes. The result is that the lines of constant H-content are crowded together on the lower part of the main sequence."

A difference will immediately be noticed between KUIPER's conclusion and the manner in which the colour-luminosity arrays of the clusters were treated above. By relying heavily on spectroscopic parallaxes—a reliance that, as he realized, is somewhat weakened by circular reasoning—KUIPER found that, although the main sequences of the clusters on the hypothesis of equal hydrogen content were not displaced from one another as much as might be implied by Fig. 14, nevertheless the spread in luminosity for different clusters at, say, a colour of P–V = $+0^m5$ might be a half a magnitude or more. If this is so, then the distances derived above by standardizing the Praesepe main sequence with the use of parallax stars, and then force-fitting the main sequences of the other cluster to it, will lead to a differential error that increases from the Pleiades to M 67; the absolute error will of course depend on the hydrogen content of the parallax stars used in the standardization.

It is, therefore, highly desirable to test the fundamental assumption made in constructing the colour luminosity arrays of the clusters, that is, that the main sequence for P–V greater than about $+0^m4$ are identical with one another and with that of the nearest stars. There is evidence to the contrary in the colour-luminosity array shown in Fig. 15 which was constructed from 350 stars, each with a mean parallax (from at least two trigonometric determinations at different observatories) greater than $0''.05$; known double stars were excluded unless the components differ by 0^m5 or less, in which case the mean colour and luminosity of the components is plotted, or if the magnitude difference exceeds 2^m5 and the companion could be disregarded.

Even if we disregard the obvious giants, the main sequence in Fig. 15 seems to have, for a given colour, a width of at least a magnitude. Of course, despite the fact that only twice observed parallaxes larger than $0''.05$ have been used, some of the spread can still be attributed to errors in M_V, but if we still further refine the data, favouring only the most heavily weighted values of the parallax, the spread persists until we reach a point where further refinement would leave too few stars for discussion, when the red end of the sequence splits into two sequences—or ridge lines—separated by about a magnitude. This latter result will be referred to again.

Unfortunately, the parallaxes of the stars in the clusters discussed above are either too small to be measured or, if measured, so small as to be subjected to uncertainties much greater than the one magnitude spread of the main sequence in Fig. 15. There is, however, a cluster not yet discussed, the Hyades, which is so close to the Sun that its depth (8 to 10 parsecs) is an appreciable fraction of its distance (40 parsecs). The parallax (0″·025) is still too small to be reliably measured trigonometrically, but the large depth to distance ratio produces easily measurable variations in

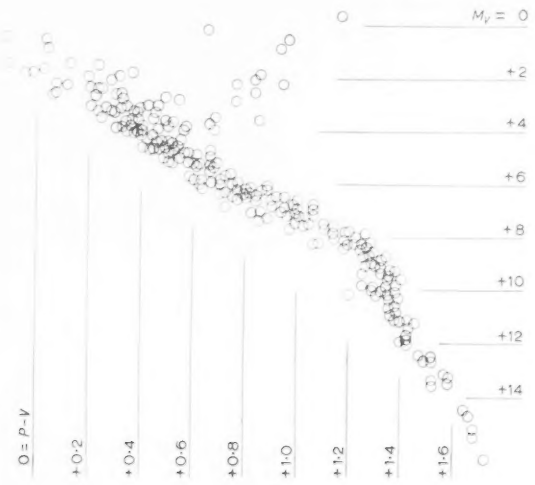


FIG. 15

the size and direction of the proper motions of the individual stars, dependent upon their location in the cluster, which permit, from the assumption that the stars are actually moving together in space, a determination of the apparent convergent point of these proper motions. The co-ordinates of this point (A, D) are:

$$A = 6^{h}18^{m}$$
$$D = +7°5$$

Furthermore, the radial velocities, ρ_0, of about 125 cluster members have also been measured, giving a mean value of $V_t = \rho_0 \sec \lambda = 44$ km/sec, where λ is the angular distance of each star from (A, D). We can then compute the parallax of each cluster member from its total annual proper motion μ and the relation $\pi = 4{\cdot}74\,\mu/V_t \sin \lambda$. This has been done for the known members and the colour luminosity array is plotted in Fig. 16.

The outline of the Praesepe colour-luminosity array from Fig. 3 is also shown in Fig. 16 and it is obvious that the main sequence of Praesepe, standardized by the nearest stars in Fig. 2, is identical with that of the Hyades, independently determined by the cluster parallaxes. Indeed, the two clusters are nearly identical in all respects; the main sequence with the sprinkling of double stars (some known, others surmised as in Fig. 1 and indicated by crosses), the one singularly blue star, the four red giants, the two-colour relationships in Fig. 17. In fact, this great similarity

between the two clusters in itself weakens any attempt to generalize on the coincidence of the main sequences and justify the force-fitting of the other clusters discussed above. The argument would be that the main sequences are identical because the clusters are identical or, following the STRÖMGREN-KUIPER hypothesis, because the hydrogen contents are equal. The plausibility of this argument is heightened when we consider the space motions of the two clusters. From the observed mean proper motion and radial velocity of the Praesepe cluster, and the distance derived above from equations (1) and (2), we find the following three vectors of the cluster's galactic motion with respect to the Sun; the same vector motions for the Hyades were based on the mean cluster parallax:

	Praesepe	Hyades
U (directed away from the galactic centre)	+40	+40 km/sec
V (in the direction of galactic rotation)	−21	−18 ,,
W (toward the galactic pole)	− 7	− 2 ,,

If we reverse the argument and derive the parallax of the Praesepe cluster from the assumption that it is moving in space with the Hyades, we obtain $\pi = 0''.0075$, which

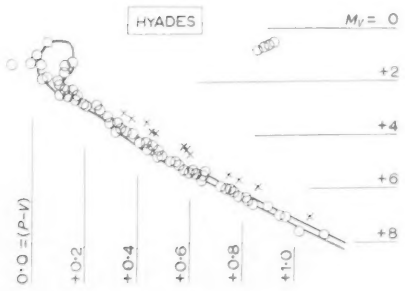


FIG. 16

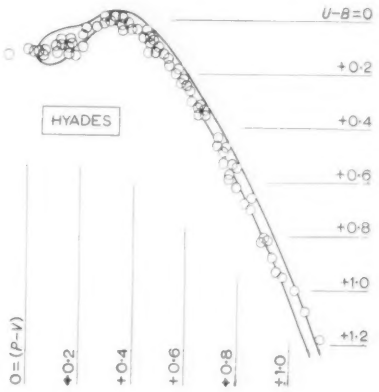


FIG. 17

leads to $U = +40$, $V = -19$ and $W = -2$ km/sec. In view of the smallness of the proper motion of Praesepe the agreement is satisfactory. The fact that Praesepe and Hyades are probably moving together in space was mooted nearly 50 years ago by HERTZSPRUNG and K. SCHWARZSCHILD; and several intervening investigations, especially one by R. E. WILSON, have shown that about 100 bright stars, twenty or so of which are within 20 parsecs of the Sun, also share this motion. Another "supercluster", or group, of stars that share the smaller, but well-determined, space motion of Sirius ($U = -14$, $V = 0$, $W = -12$ km/sec) was also noted by HERTZSPRUNG and others. The 50 odd bright stars, culled as the most probable members from several lists of possible members of the group, were used to construct the colour luminosity array shown in Fig. 18. The relationship between the P–V and U–B colour for these stars is also compared with the standard, the Praesepe–Pleiades relationship in Fig. 19.

The group membership of these stars is decided by the fact that their apparent motions, that is, their radial velocities and proper motions, appear to converge to the same point (A, D) toward which Sirius is moving, $A = 20^{h}\ 44^{m}$, $D = -42°7$, with a space velocity relative to the Sun of $V_t = 18{\cdot}4$ km/sec. The parallaxes are then computed, as for the Hyades stars above, from the relation $\pi = 4{\cdot}74\,\mu/V_t \sin\lambda$.

In addition to the colour-luminosity array for the Sirius Group, Fig. 18 also contains, in outline, that of the Pleiades from Fig. 7. The main sequences up to P–V

Fig. 18

near $-0^{m}2$ are apparently nearly identical and by comparison with Fig. 9 we would place the Sirius Group between the Pleiades and the Coma Berenices cluster in the progression of break-away points illustrated in Fig. 13. The Sirius Goup appears to contain many more giants than the Coma Berenices cluster but this probably only results from the greater total population. The large number of *intrinsically* bright main sequence stars in the Sirius Group, compared to the number fainter than, say,

Fig. 19

$M_V = +4^{m}$, is an effect of observational selection since the group members were selected from the *apparently* bright stars for which relatively more accurate proper motions and radial velocities are available.

To complete the argument that the force-fit method of determining the distances of clusters, used above, is valid, we will mention two additional groups of stars—the ζ Herculis Group and the 61 Cygni Group.

The mean space motion of the Sun, with respect to the observed stars in the *general field* is, in the co-ordinates already described, close to $U=+11$, $V=-17$ and $W=-7$ km/sec, that is, these vectors (with the opposite sign) represent the average space motion relative to the Sun of the observed stars. The average total space motion of these stars is then 21·4 km/sec and the value of $V_t = 18{\cdot}5$ km/sec for the Sirius Group is near the average, whereas $V_t = 44$ km/sec for Hyades-Praesepe is relatively large. There are a few stars for which the well-determined space motions are still higher, 75–100 km/sec; three examples, all with accurately known proper motions, radial velocities and distances, are the following:

Star	U	V	W	V_t	
β Hydri	+53	−43	−28	74	km/sec
ζ Herculis	+54	−47	−24	74	,,
61 Cygni	+92	−53	− 6	108	,,

The first two stars are apparently moving together, toward $A = 5^h48^m$ and $D = -16°5$, and a search through the stars whose well-determined proper motions and radial velocities are known to be large, but whose trigonometric parallaxes are more

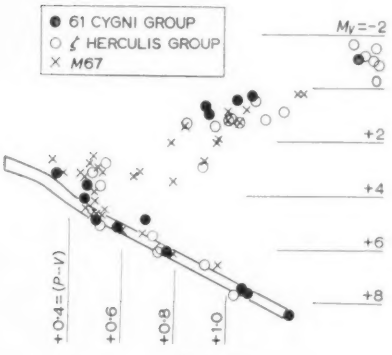


FIG. 20

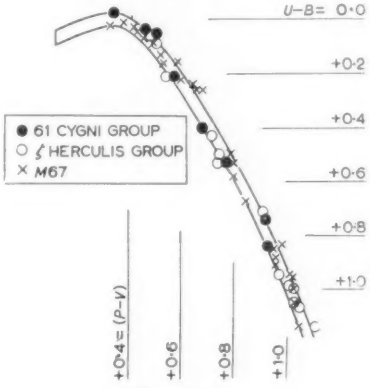


FIG. 21

uncertain, revealed 22 additional stars apparently moving towards the same point. The parallaxes of these stars have been computed on the assumption that they appear to move toward this convergent point because they actually possess the same space motion as ζ Herculis, and the resulting colour-luminosity array is shown in Fig. 20. A similar search for stars that may share the space motion of 61 Cygni revealed 14 possibilities whose colour-luminosity array is also shown in Fig. 20. The members of M 67 in Fig. 13 whose colours are similar to the stars represented in Fig. 20 are replotted in the latter figure. Fig. 21 shows the relationship between the two colours for the Group and for M 67 stars, compared to the standard Praesepe relationship of Fig. 4, and there is no reason to suspect that interstellar absorption has affected the observations; it must be kept in mind that the determination of U–B is, compared to P–V, a relatively difficult observation, especially for the faint stars in M 67.

If we accept the evidence offered by the stellar groups—the nature of which is not completely understood, nor is their existence unequivocally established and universally accepted—the force-fit method of determining the cluster distances seems to be

justified for Praesepe by the Hyades, for the Pleiades and Coma Berenices by the Sirius Group and for M 67, and, perhaps, for NGC 752, by the ζ Herculis and 61 Cygni Groups. Or, to put it another way, the main sequences, for P–V redder than about $+0^{m}4$ or $+0^{m}5$, seem to be identical in all the clusters and groups discussed.

If the main sequences are as similar as they appear to be for the clusters and groups discussed above, then the large spread in that of the colour-luminosity array for the nearer stars, in Fig. 15, needs an explanation. Taking a clue from the above discussion of groups, we will add another parameter to Fig. 15; the space motion. The proper motion, radial velocity and trigonometric parallax of each of the 350 stars have been used to compute the vector motions U, V, and W, which, in turn, define the individual galactic orbits. These orbits were then divided into classes, in the following manner, suggested by R.v.d.R. Woolley:

Class	Y
A	0·5
B	0·5–1·5
C	1·5–3·0
D	3·0–5·0
E	5·0–6·5
E+	6·5

where Y is the distance from the Sun's galactic orbit to that point in the star's orbit which is closest to the galactic centre (pericentric distance). Stars of Class D, for example, pass within 7 kiloparsecs of the galactic centre and those in Class E+ within 3·5 kiloparsecs. The absolute values of Y are uncertain since they depend upon an assumed distance from the Sun to the galactic centre (10 kiloparsecs) and upon adopted values for the so-called constants of galactic rotation (Oort constants) but regardless of numerical size, the classification from Class A to Class E+ represents a gradation of increasingly close approaches to the centre of the Galaxy.

The colour-luminosity arrays for the stars in classes A to E+ are shown in Figs. 22–27, where a section of the Pleiades main sequence is also shown in outline.

The source of much of the spread in the main sequence of Fig. 15 is obvious in Figs. 22–27. In the first place, each class of stars presents a colour-luminosity array that looks very much like that for a group or cluster, with a regular progression

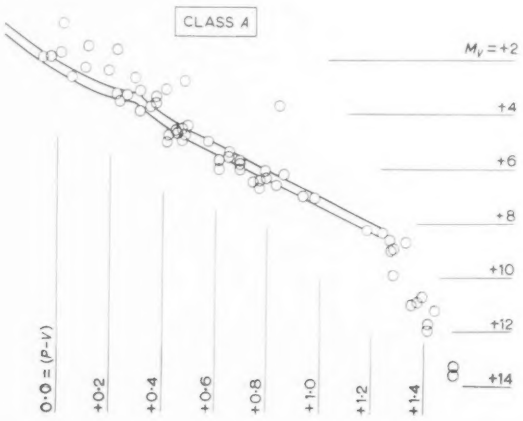


Fig. 22

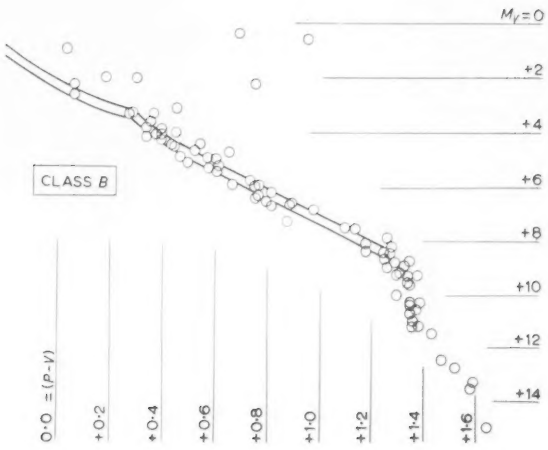


FIG. 23

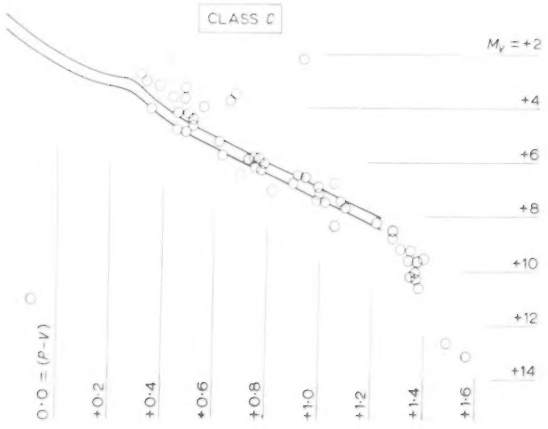


FIG. 24

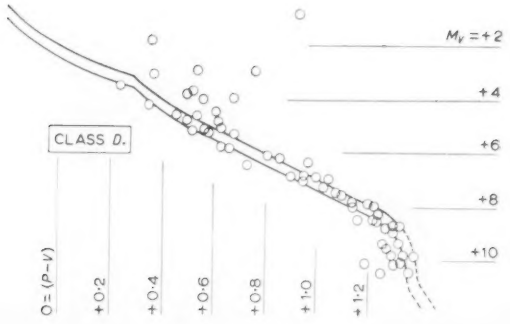


FIG. 25

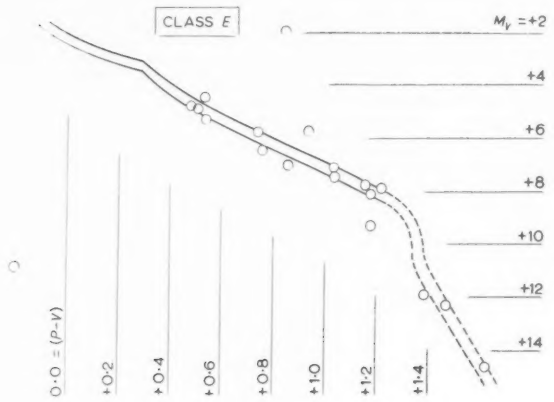


FIG. 26

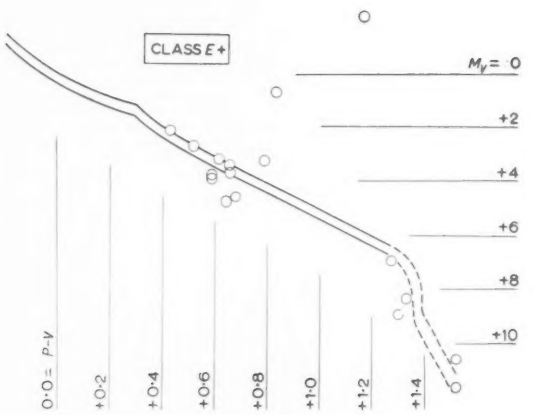


FIG. 27

toward increasing values of P–V of the break-away points as we go from Class A to Class E+. Since each break-away point creates a scatter above the main sequence, the combined colour-luminosity array of Fig. 15, at least to P–V ~ $+0^m5$, is broadened greatly. The same effect would appear if we plotted colour-luminosity arrays for all the clusters and groups discussed above in a single diagram. Secondly, it appears that, in spite of the evidence of the stellar groups, there may be at least two main sequences for stars redder than P–V ~ +0·5. A few stars tend to broaden the main sequence of the Class A stars in Fig. 22 and Class B in Fig. 23, but they might easily be explained away as poor parallaxes. However, easily visible in Class C (Fig. 24) and obvious in Class E+ (Fig. 27) are definite subdwarfs, forming a sequence about three-quarters of a magnitude below the main sequence, and parallel to it.

Because the picture is blurred by the inevitable presence of uncertain parallaxes, we are unable to judge from Figs. 22–27 whether or not the dwarf and subdwarf

sequences are discrete. There are observations, however, that will ultimately be available and which may answer the question in the following way. Although we presented a two-colour plot of U–B against P–V for the members of the clusters and groups discussed above, this was not done for the objects represented in Fig. 15 because values of U–B are available for less than half the stars and some of those that are available, are of low accuracy because of the extreme faintness of many of the stars. The values that are available, however, indicate that not only is displacement below the main sequence connected with an ultra-violet excess, that is, with a smaller than normal value of U–B, but that the amount of ultra-violet excess may be an indication of the amount of displacement. It should be mentioned in this connection, that a few values of U–B indicate that the clump of a half-dozen M-type stars displaced from the main sequence of Class C (Fig. 24) near P–V = $+1^m2$ to $+1^m3$ have a strong ultra-violet excess.

The grouping of the stars into Classes A to E+ depends upon both the orientation and the eccentricity of the galactic orbits, but, in general, the greater the eccentricity the greater the space velocity and the later the class. The colour-luminosity array for the high-velocity groups 61 Cygni and ζ Herculis in Fig. 20 is very similar to that for Class D in Fig. 25 and, indeed, contains several of the same stars. The colour-luminosity arrays for the Sirius Group (Fig. 18)—Class A—and Hyades (Fig. 16)—Class B—are quite similar to those in Figs. 22 and 23, respectively. It appears to have been an accident that the 61 Cygni and ζ Herculis Groups, used to justify the main sequence of M 67, have this sequence identical with Praesepe, because a different choice of stars, a group containing subdwarfs, would have led to a different conclusion. In fact, there is a group of about a dozen Class E stars that share the space motion of the well-known subdwarf μ Cassiopeiae. In this direction, however, we have reached the borders of what a summary report can contain and, in a sense, we have come full circle, for, although perhaps not in the clusters discussed by KUIPER or for the exact reasons suggested by STRÖMGREN, we are not left with an unambiguous definition of a main sequence.

We will retreat to 1933 when STRÖMGREN, in the classical paper already referred to, said, "One would . . . expect that the tracks of evolution in the H–R diagram are the lines of constant mass (broken lines in Fig. 14). One would further expect that the hydrogen content decreases so that the stars expand". This line of reasoning has been developed by several investigators, including CHANDRASEKHAR, SANDAGE, SCHWARZSCHILD and HOYLE. In a rough, and extremely simplified way, the argument is as follows. There is a main sequence populated by homogeneous stars consisting of a convective (highly turbulent) core and a radiative (relatively quiescent) envelope; the core contains 10–15 per cent of the mass of the star. If the star is similar to the Sun, then, through the carbon–nitrogen cycle or, perhaps, the proton–proton chain, first the hydrogen in the core and then that in a shell of increasing size around the core is converted into helium, causing the luminosity to rise and placing the star approximately a magnitude above the original main sequence with an essentially unchanged surface temperature or colour. The core then begins to contract, releasing energy in much the same way as that proposed by HELMHOLTZ in the quotation on page 258, and causing a rapid expansion of the envelope and therefore a decrease in the surface temperature, that is, the star moves nearly horizontally

to the right in the colour-luminosity array. On the other hand, if the star is of B-type, P–V $\sim -0^m2$, its structure may be radically different and the consumption of hydrogen more rapid, causing the star to move nearly directly from left to right in the colour-luminosity array. The rate at which these processes take place is proportional to the ratio of luminosity to mass, and therefore increases with luminosity, so that at any given time the locus of points representing stars in a coeval group, or cluster, will in fact look somewhat like the lines of equal hydrogen content in Fig. 14; the evolutionary paths of the individual stars will be similar to the broken lines of equal mass in the figure. However, the differential rate of evolution is such that practically no star redder than P–V $= +0^m5$ has evolved at all and, consequently, the main sequences for all stars come together at this point.

Therefore, with suitable assumptions, the progression of break-away points illustrated for the clusters in Fig. 13 can be interpreted to indicate the relative ages of these clusters, since the age is a function of the luminosities of those stars that are just beginning to undergo a change of structure due to hydrogen consumption and therefore just moving off the main sequence.

The Pleiades cluster (Fig. 7), for example, is a young cluster and its main sequence is—or is close to—the primeval sequence, whereas the Sirius Group (Fig. 18) is somewhat older and shows its age by the existence of several red giants and the sharply defined break-away point near $M_V = 0$. The B-type stars, once the brightest stars in the group, have burned out their hydrogen, appeared perhaps briefly as red giants, and passed on to what is yet an uncertain fate. Some of the early A-type stars have moved rapidly to the right and now are the yellow giants near $M_V = 0^m$, P–V $= +0^m8$, while others have only begun this change and appear as A-type giants with P–V between -0^m1 and $+0^m1$; the exact present location of the original A-type main sequence stars, on the $M_V \sim 0^m$ path from P–V $= -0^m1$ to $+1^m6$, perhaps depends on such individual characteristics as the rate of rotation, which may control the availability of the hydrogen and therefore the rate of evolution. The presence of gaps in the distribution along this $M_V \sim 0^m$ track—such as the "HERTZSPRUNG gap" between P–V $= -0^m1$ and $+0^m7$ in Fig. 18—may indicate that the course of evolution is not smooth and the snapshot represented by the colour-luminosity array is less liable to catch stars in one stage of the process than in another.

Proceeding to NGC 752 (Fig. 11), although a few of the original main sequence F-type stars linger as G-giants, all the A- and B-types have disappeared into the Never-Never along tracks that, at the present, cannot even be described as esoteric. And in M 67 (Fig. 13), with the exception of a few evolutionary laggards, still lingering on or near the early-type main sequence, only the dwarfs of G-type or later are left. The range, from $M_V = +3^m5$ to 0^m, of the giants in this cluster, and in the 61 Cygni and ζ Herculis Groups (Fig. 20), is larger than that of the other clusters because the evolution in this very old cluster has reached stars of relatively low luminosity and, since its rate is related to that luminosity, the ultimate and still unknown fate which will remove the stars from this section of the colour-luminosity array is longer postponed.

The colour-luminosity arrays of the clusters and groups, then, differ only because the clusters are in different stages of their development. How can we test this conclusion?

We assume that, in general, a star's mass is conserved in its evolution, which means that a star in Fig. 18, for example, now at $M_V = 0^m$, and P–V $= +0^m8$ still is as massive as it was when at, say, $M_V = +1^m$, andP–V $= -0^m1$. Unfortunately none of the stars in Fig. 18 near these points are visual or eclipsing binaries, so a mass determination is not possible. However, there are two visual binary systems for which well determined orbital elements and parallaxes are available and which fall in about the designated points in the array—α Aurigae and ζ Ursae Majoris. The colours and luminosities of the individual components of α Aurigae and the mean of the equal components of ζ Ursae Majoris are plotted in Fig. 28, where the area enclosed by continuous lines indicates the run of the colour-luminosity array for Coma Berenices from Fig. 9. The orbits and parallaxes of these systems give, in terms of solar mass, 3·0 ⊙ for α Aurigae and 3·1 ⊙ for ζ Ursae Majoris—in each case the mean of both components. The bright components of the visual binary α Ursae Majoris and the eclipsing system β Persei, which are also plotted in Fig. 28, have well-determined masses of 6·8 ⊙ and 5·0 ⊙, respectively. Allowing for the inevitable

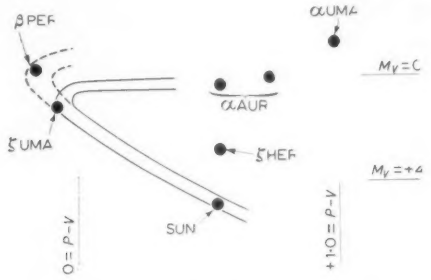


FIG. 28

uncertainty in even such well-determined masses as these, they do not contradict the assumption that α Aurigae and α Ursae Majoris have evolved from stars similar to ζ Ursae Majoris and β Persei. Also, the mass of ζ Herculis, 1·1 ⊙, compared with that of the Sun is consistent with the possibility discussed above, that it evolved from the main sequence, somewhat bluer than the Sun, and has brightened up about a magnitude in the process. Unfortunately, many of the available mass determinations are for components of eclipsing systems, and the fact that a large percentage of these consist of stars close enough together to have affected each other's evolutionary processes, makes the resulting conclusions inapplicable to single stars.

The distribution of the giants in the colour-luminosity array for M 67 (Fig. 13) is of particular interest. As already mentioned, the large range in luminosity of the giants may result from the fact that the evolution in this very old cluster has reached stars of relatively low luminosity and that, since the rate of evolution is related to the luminosity, the low-luminosity giants may linger longer. This, however, is not the usual explanation. In the rough outline given above of the evolutionary process, it was stated that a main sequence star similar to the Sun, after brightening a magnitude or so without reddening, would move to the right of the colour-luminosity array as a result of lowering surface temperature, caused, partly at least, by the contraction of the stellar core. This contraction, however, probably cannot maintain a temperature gradient, so that the core becomes isothermal and the stars' energy is supplied by

hydrogen burning outside the core. A hydrogen convection zone is created which steadily grows in size and the star rapidly increases its luminosity with time. This explanation for the giants in M 67, which then would all have originated on the main sequence at about the same luminosity $M_V \sim +4^m$ (although not a logical extension of the argument in terms of evolution along nearly horizontal tracks used for the other galactic clusters) is introduced because of the analogy between the colour-luminosity array for this cluster and that observed for the globular clusters, discussed below.

If the giants in M 67 have evolved in this way, then those near $M_V = 0^m$ should have near solar masses. Also, if the stars in the general field, such as those represented in Fig. 15, represent a mixture of ages, then we might expect the yellow giants to be a mixture of stars that have evolved directly from the main sequence without much change in visual luminosity and therefore have masses of 3 to 6 solar masses, and old stars that have evolved in the manner just described from the main sequence near the Sun and with, thereafter, near solar masses. Although SANDAGE has estimated that the older stars of solar mass should outnumber the younger ones by 3 to 1, the only reliably known masses for yellow giants near $M_V = 0^m$ are comparable with those of main sequence stars of the same luminosity—but the number is small.

The fact that stars can arrive at $M_V \sim 0^m$ to $+2^m$ as young stars, moving nearly straight across the colour-luminosity array, and as very old stars, moving up from the main sequence near the Sun, gives rise, by means of what SANDAGE has called the "funnel-effect", to a heavy concentration of K0–K2 giants and may explain the observed fact that the luminosity function of K-type stars in the general field shows a maximum near $M_V = +1^m$. Also called in evidence, in support of the funnel-effect, is the fact that the velocity dispersions of the G–K giants and G dwarfs are similar and are both greater than that of the A-type stars.

Another way of looking at this latter effect, for the nearer stars, is shown in Fig. 29, where the colour-luminosity arrays for all stars apparently brighter than visual magnitude 5 and absolutely brighter than $M_V = +4^m$, whose parallaxes determined by at least two observatories are greater than $0''.02$, are plotted—the sorting of the stars into classes A to E+ was carried out in the manner described above for the stars within 20 parsecs of the Sun. The main sequences beyond about P–V $= +0^m5$ are not represented, but, supposedly, they would be nearly the same for all classes as indicated in Figs. 22–27.

Since the classifications are based on the eccentricity of the galactic orbits and, therefore, in general on the velocities, both the early and late type stars in Class A, for example, have the same dispersion in velocity. If we combine the Class A stars with, say, those in Classes D and E, the velocity dispersion of the combined group of late type giants will be increased, but that of the early types—since Classes D and E contain none—will be unchanged. Therefore, any group of stars chosen from the general field will show a greater velocity dispersion in the late-type giants than in the A-type stars, but no conclusions about the origin of these giants can be based on this fact alone, since the velocity dispersions of the G dwarfs of any one class, A to E+, will be identical with that of the giants, as well as of the A-type stars, in the same class.

A further source of information, bearing directly on stellar evolution, is available in the colour-luminosity arrays for globular clusters. Because of the faintness of the

stars in these clusters and, therefore, the difficulty of observations, the exploitation of this field is only beginning. Preliminary studies of several clusters have been published; the most extensive material available is that for M 3, whose colour luminosity array is shown in Fig. 30; the outlined array is that of the Pleiades. The plotted points represent mean values of the visual magnitude in groups of ten or more stars, converted to absolute magnitudes by fitting the faintest stars to the Pleiades main sequence. The shape of the subgiant–giant sequence, B, is seen to somewhat resemble that of M 67 (Fig. 13) and, also similar to M 67, there is a blue

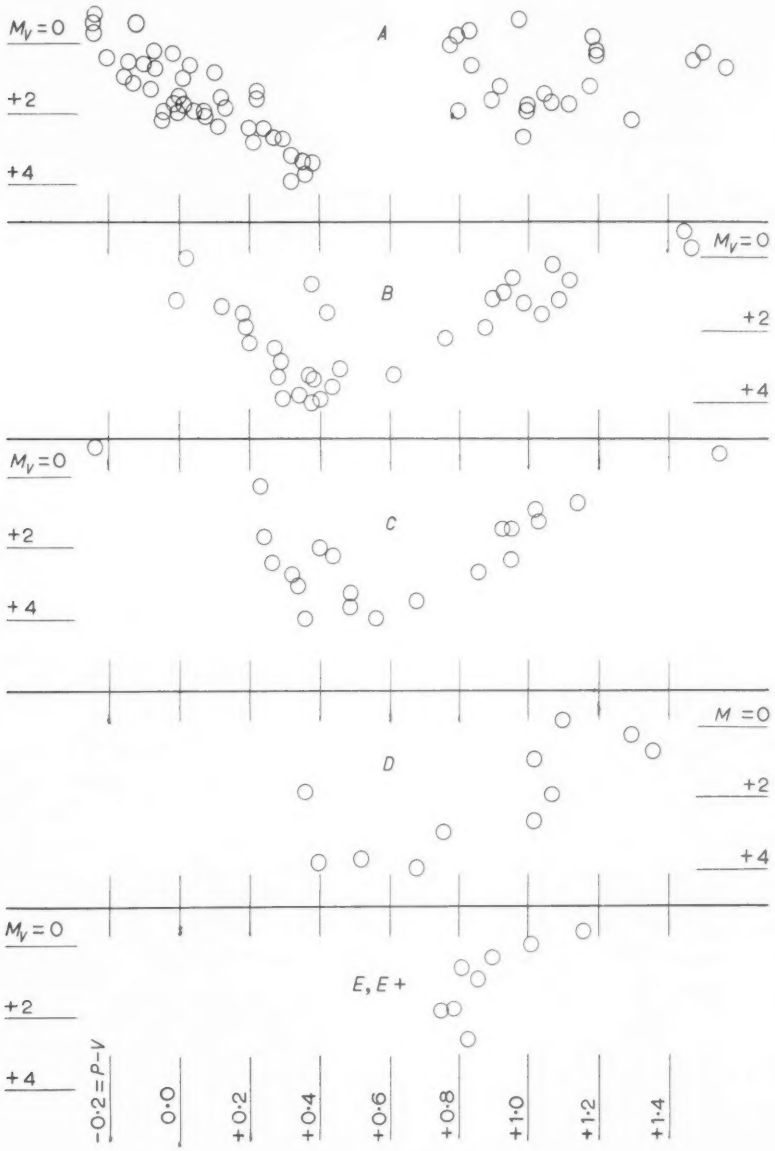


FIG. 29

horizontal sequence, C. The RR Lyrae or cluster-type variables in M 3, and probably in all globular clusters, fall in a well-defined region of the horizontal branch; this region for M 3 is indicated by a bar in the centre of sequence C. On rather shaky, but the best available, evidence, these variables are usually regarded as having $M_V = 0^m$, and the value of $+0^m2$ derived for those in M 3, by fitting the main sequences in Fig. 30, is not unreasonable. The two-colour diagram for the mean points in Fig. 30 is shown in Fig. 31, where the location of the stars that populate the sequences A, B and C is also indicated; the outlined area is that defined by the Praesepe–Pleiades two-colour relationship. The stars of sequence B deviate noticeably from the standard relationship, and those of sequence A deviate by as much as half a magnitude. The stars of sequence A are, of course, the faintest observed in the cluster, reaching magnitude 21, and the U–B colour, a difficult observation to make under much better conditions, becomes suspect when obtained for such faint objects in the crowded field of a globular cluster. However, although the exact position of

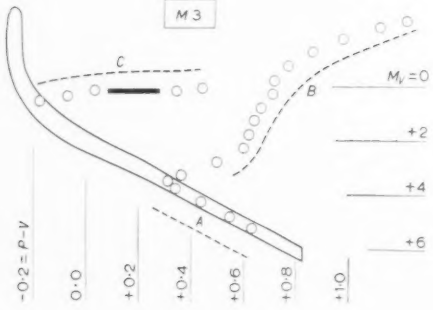


Fig. 30

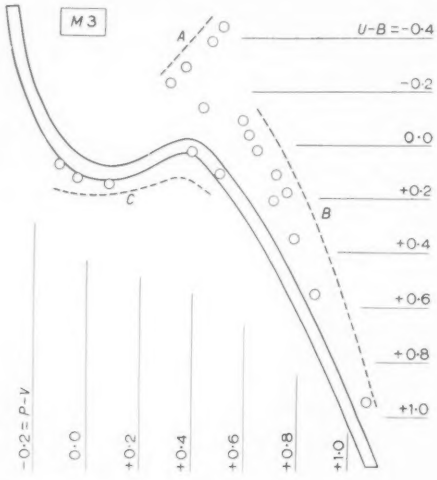


Fig. 31

sequence A stars in Fig. 31 may be in some doubt, the run of sequence B stars above the standard relationship—that is, the stars show an ultra-violet excess—is certainly real and makes impossible a determination of the reddening of the cluster by the method indicated above for the Pleiades. This is especially true when the position of sequence C stars in Fig. 31 is considered, since they show, if any deviation at all, an ultra-violet deficiency. A reddening correction might be determined from the sequence C stars, but if the sequence A and sequence B objects are not similar to the standard Praesepe–Pleiades stars, there is no reason to think those on sequence C are more normal.

The outlines of the colour-luminosity array and the two-colour relationships for M 3 are repeated in Figs. 32 and 33, where a synthesis of several preliminary investigations of the globular cluster M 13 is also shown schematically. The luminosities for the M 13 stars were obtained by fitting the faintest cluster stars to the Pleiades main sequence. The two-colour relationship in Fig. 33 is something of a surprise, since it indicates no ultra-violet excess for the sequence B stars; U–B colours for

sequence A stars in M 13 are not yet available. There were only indications of an ultra-violet deficiency in the sequence C stars of M 3, but this deficiency is undeniable in M 13. Sequence C for M 13 in Fig. 32 is somewhat idealized, especially in the region occupied by cluster-type variables indicated by the unshaded area, since this cluster is very poor in such variables compared to M 3.

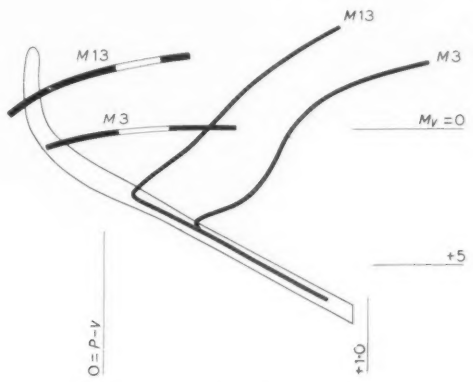


FIG. 32

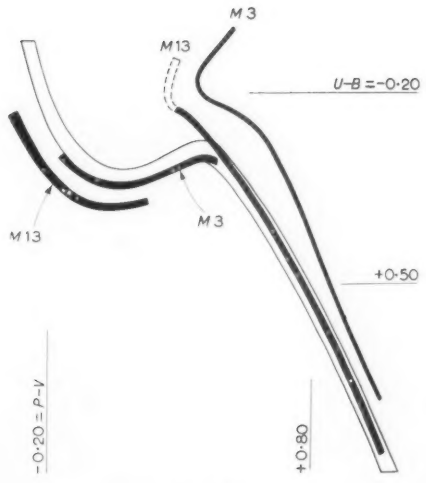


FIG. 33

Whereas force-fitting the main sequence of M 3 to that of the Pleiades gave a value of the luminosity of the cluster-variables reasonably close to that expected, the same procedure for M 13 puts those in that cluster near $M_V = -2^m$. If we reverse the procedure, and assume the variables are near $M_V = 0^m$, the main sequence of M 13 falls 2 magnitudes below that of the Pleiades and is reminiscent of the subdwarfs among the Class C, D and E stars near the Sun in Figs. 24–27. Also, like the globular cluster stars, these later subdwarfs show an ultra-violet excess. The material available for other globular clusters indicates that the problem is widespread and that there is

either a series of main sequences for globular clusters or a series of choices for the luminosity of the variables, depending upon which we choose to fix. This quandary is a barrier both to the interpretation of the arrays for globular clusters and their use in distance determinations until we can observe several varieties of clusters at some fixed distance—such as those in the Magellanic Clouds—and learn their class characteristics. Even a study of the clusters in the Magellanic Clouds may give us information that is not transferable to field clusters, but, from an evolutionary standpoint, they are of enough interest in themselves.

An interesting question, in connection with the colour-luminosity arrays of globular clusters, is the origin of the RR Lyrae stars. If the stars in such clusters are as old as we believe, any objects originally on the main sequence at $M_V = 0^m$ that have as yet only reached the RR Lyrae region must either have had a stunted evolution because of, perhaps, an abnormal constitution, or have arrived at that region in its normal evolution from some other part of the array. The usual argument is that the latter situation is true and that the stars that are now cluster-type variables were originally main sequence stars of solar mass—a plausible argument or one that can be made plausible with the right assumptions. A very sketchy outline of this argument is, that, when the stars have reached a certain point (near $M_V = -3^m$) on their course from their origin as solar type stars on the main sequence through subgiants to bright red giants, because of mass loss or perhaps a change to helium burning, the stellar conditions are suddenly altered and, as a result, the star's luminosity is suddenly lowered. This stage may be followed by some form of general "collapse" which moves the star into the cluster-type variable region of sequence C (Fig. 30) and ultimately to the white dwarf region near $M_V \sim +11$ and P–V $\sim -0^m3$ to $+0^m5$.

The white dwarfs, which have not been previously mentioned, also occur in a few galactic clusters, such as the Hyades and, perhaps, Praesepe and Coma Berenices, and are thought to be the remains of the early-type stars in these clusters, which, presumably, arrived at the white dwarf stage through a different process than that just described, since the existence of a theoretical upper limit to the mass of a white dwarf requires some form of fragmentation, e.g. a nova explosion. But these stars are necessarily faint and our knowledge of them is so incomplete that nothing can be done at present to equate their numbers to those expected in galactic clusters.

"In listening to this discussion," said Sir Ernest Rutherford during a symposium held at the Manchester meeting of the British Association in 1915, "it seemed to me that astronomers may be proceeding too much on the assumption that the evolution only proceeds in one direction. That does not seem to be necessary. I see no reason why we should not have some stars condensing and others diverging." Some recent work by Henyey, Herbig and Walker, summarized below, bears testimony to Sir Ernest's perspicacity.

The youngest, on the evolutionary hypothesis, of the clusters so far discussed is the Pleiades, which is also the only cluster that appears to be appreciably affected by interstellar absorption. The well-known connection between the galactic distributions of absorbing matter and of early-type stars would lead us to expect that still younger clusters, that is, those containing O-type stars, would be still more affected. Such a

cluster is NGC 2264 for which the two-colour relationship, from photoelectric observations of probable members, is shown in Fig. 34; the standard Praesepe-Pleiades relationship is also shown in outline.

A striking feature of Fig. 34 is the well-defined two-colour relationship for stars with P–V bluer than 0, which is displaced from the standard relationship in a direction and by an amount that indicates a reddening of $E_{P-V} = +0^{m}08$. The relationship for the redder stars, however, is nearly non-existent. The cluster is seen projected against an almost completely opaque absorbing cloud, so that the possibility of explaining the position of the redder stars in Fig. 34 as due to very large amounts of reddening of O- and B-type stars seems untenable in view of the apparently uniform reddening of the bluest stars, unless, of course, the background stars

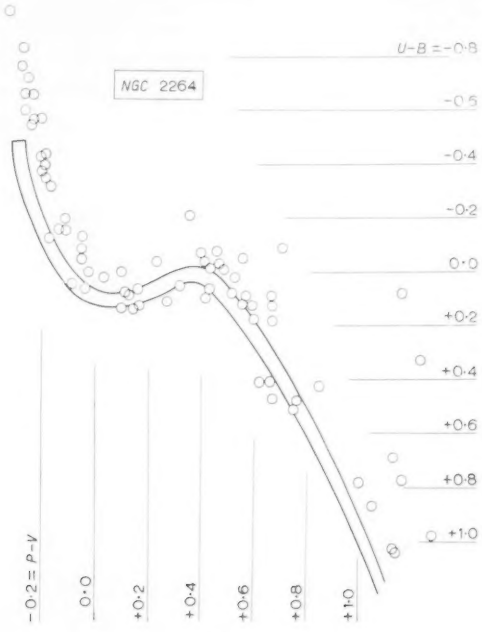


FIG. 34

themselves formed a cluster imbedded in the near side of the cloud. However other evidence, including spectroscopic observations, show that some of the red stars are not of early type and it seems well established that they, like the stars in some globular clusters, show an ultra-violet excess compared to the standard Praesepe stars of the same P–V.

Ignoring, for the moment, the impossibility of determining the absorption from the redder stars and adopting the value $E_{P-V} = +0^{m}08$ derived from those bluer than P–V $= 0^{m}0$, we have fitted the faintest stars with corrected colour redder than P–V $= +0^{m}5$ to the standard main sequence, and derived a distance modulus of $V_0 - M_V = V - 3E - M_V = +8^{m}2$. The resulting colour-luminosity array is shown in Fig. 35, where that of the Pleiades is also shown in outline. There are several surprising features in Fig. 35. The position of the main sequence, for P–V bluer than $0^{m}0$, might have been anticipated for a cluster that is younger than the Pleiades;

but the presence of yellow giants, four or five magnitudes *fainter* than the brightest early-type stars, and the large spread in the main sequence, are features not met with in the arrays for the galactic clusters already discussed. There are more stars, redder than P–V = 0^m0, above the main sequence than on it, and the spread is larger than can be explained by the presence of double stars. Also, there may be a correlation between the ultra-violet excess, shown in Fig. 34, and the distance *above* the main sequence in Fig. 35—in contrast to the suspected correlation for the nearby subdwarfs in Figs. 24–27, and perhaps for the globular cluster stars, between this excess and displacement *below* the main sequence. Additional evidence is necessary not only to definitely establish these correlations but also to disentangle them, for which purpose recourse to wavelength regions other than that used for the U-magnitudes may be necessary.

The method used in deriving the colour-luminosity array for NGC 2264 in Fig. 35, that is, by fitting the fainter stars to the standard main sequence, is not that usually

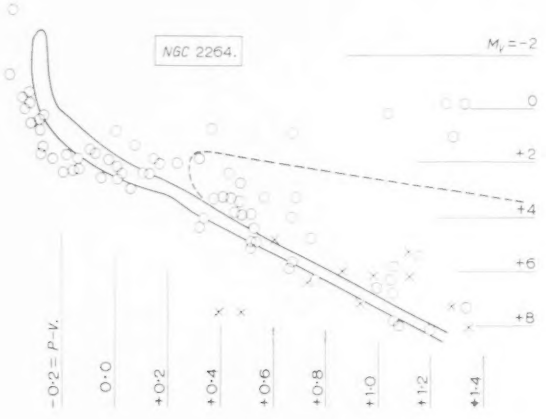


FIG. 35

adopted for such young clusters. Instead, it has been assumed that only those stars bluer than P–V ~ 0 are main sequence objects, and they have been fitted to an unevolved or "zero-age" sequence—similar to that defined by the early-type stars of the Pleiades, but approximately 0^m5 fainter to allow for the evolution of these latter stars. When this is done, even the faintest members of NGC 2264 redder than P–V ~ 0 fall 1^m5 to 2^m0 above the standard main sequence, a position that is justified by the limited spectroscopic evidence that some, at least, of these stars are subgiants.

The most interesting feature of these very young clusters is the fact that they contain representatives of the so-called T Tauri variables. The colours and luminosities of some of these erratic variables are shown as crosses in Fig. 35—the colours and luminosities are instantaneous values, since the mean or the extreme values are not known. No attempt has been made to indicate the position of these stars in Fig. 34, since the values of U–B occasionally change by as much as one magnitude for only one or two tenths of a magnitude change in V or P–V, possibly because of

intense emission. It is interesting that the spectra of these T Tauri stars, as well as some of the non-variable G–K stars in NGC 2264, show broadened lines which, if interpreted as resulting from large rotational velocities of the stars, is a result in direct contrast to the fact that for G–K stars in the other galactic clusters discussed, and in the general field, rotationally broadened spectral lines are extremely rare. The fact that such line-broadening is, in the late spectral types, mostly confined to the W Ursae Majoris type eclipsing binaries might indicate that the two types of variation are due to similar physical conditions—especially since large and erratic variations in the ultra-violet are comon to both—but whereas the W Ursae Majoris variables are mainly double-lined spectroscopic binaries, the T Tauri stars are not, so that the similarities are at best indicative of only a generic relationship.

Clusters such as the Pleiades and M 67 show their age by the fact that the early type stars have moved to the right in the colour-luminosity array, but the interpretation placed on NGC 2264 and similar clusters is that they are so young that the stars, which will eventually end up on the main sequence, are still contracting out of the primeval gas cloud. Since the rate of evolution in this, as in later stages, is directly related to the luminosity, the early-type stars have already reached the main sequence, but most of the later-type dwarfs—and perhaps, in NGC 2264, a few of the red giants in Fig. 35—are still on the way. Pure gravitational contraction carries the star across the colour-luminosity array, approximately along a track similar to that indicated in Fig. 35 by a broken line, with only a slight increase in luminosity, but with rapid increase in surface temperature (colour) as the radius decreases. Somewhat before the proto-star reaches the main sequence, nuclear reactions begin and contraction gradually ceases, causing a decline in the luminosity and the beginning of a logarithmic approach to the main sequence.

Whatever final interpretation is made of the colour-luminosity arrays of such clusters as NGC 2264, it is obvious that their stellar content—T Tauri stars, peculiar spectra, rotationally broadened lines in late-type spectra—insures that they are an important link in any evolutionary hypothesis and deserve the extensive study they undoubtedly will receive. In the meantime, examples of clusters younger than NGC 2264 or, although theoretical considerations indicate they may not exist, older than M 67 are more desirable than discussion of another Hyades or another Pleiades. In fact, further work on the Hyades or Pleiades themselves should not be overlooked—a nose count, to the limit of existing telescopes, is still not possible in a single cluster.

There are numerous other approaches—luminosity function, rotational velocity, chemical abundances—to the evolutionary question that have not been touched upon here. A few selected, mostly summary discussions of these approaches are included in the appended bibliography which also contains most of the observational contributions (photoelectric) to the question of colour-luminosity arrays, published in the last decade.

REFERENCES

ARP, H. C. . . . . 1955 Colour-magnitude diagrams for seven globular clusters, *Astron. J.*, **60**, 317.

ARP, H. C., BAUM, W. and SANDAGE, A. R. 1953 Colour-magnitude diagrams for M 92", *Astron. J.* **58**, 4.

BAPPU, M. K. V. . . . . 1954 Photometry of the Perseus Cluster, *M.N.* **114**, 687.

Baum, W.	1954	Photoelectric observations in M 13, *Astron. J.*, **59,** 422.
Bergh, S. van den	1957	Luminosity functions of Population I, *Ap. J.*, **125,** 445.
	1957	Luminosity function of M 67, *Astron. J.*, **62,** 100.
Blanco, V.	1957	Band-width effects in photoelectric photometry, *Ap. J.*, **125,** 209.
Blaauw, A.	1956	Luminosities, ages, kinematics and duplicity of the early-type stars, *Publ. Astron. Soc. Pacific*, **68,** 495.
Bondi, H.	1950	On the interpretation of the H–R diagram, *M.N.* **110,** 595.
Bonsack, W. K., Greenstein, J. L., Mathis, J. S., Melbourne, W. G., Neugebauer, G., Newburn, R. L., Olsen, K. H., Tifft, W. G., Wahlquist, H. D. & Wallerstein, G.	1957	Interpretation of the photoelectric colours for stars of types B–F, *Ap. J.*, **125,** 139.
Brown, A.	1951	Colour-magnitude array for stars in M 15, *Ap. J.*, **113,** 344.
Brownlee, R. and Cox, A. N.	1953	Magnitudes and colours in NGC 6231, *Ap. J.*, **118,** 165.
Chandrasekhar, S. and Schönberg, M.	1942	Evolution of main sequence stars, *Ap. J.*, **96,** 161.
Cox, A. N.	1954	Study of the galactic cluster NGC 2287, *Ap. J.*, **119,** 188.
	1955	Study of the galactic cluster NGC 2516, *Ap. J.*, **121,** 628.
Crawford, J. A.	1955	Subgiant components of eclipsing binary systems *Ap. J.*, **121,** 71.
Deutsch, A.	1956	The dead stars of population I, *Publ. Astron. Soc. Pacific*, **68,** 3P8.
Eggen, O. J.	1950	Colour-luminosity arrays for the Hyades Pleiades, Coma Berenices and Ursa Major clusters *Ap. J.*, **111,** 65, 81, 414.
	1950	Colour-luminosity array for stars near the Sun, *Ap. J.*, **112,** 141.
	1951	Colour-luminosity arrays for stars in Praesepe and in M 39, and the colour spectral-type relationship, *Ap. J.*, **113,** 657, 663.
	1955	Masses of the subgiants, *Publ. Astron. Soc. Pacific*, **67,** 315.
	1955	Magnitudes and colours of 833 stars and a discussion of the colour-luminosity arrays for parallax and cluster stars, *Astron. J.*, **60,** 65, 401, 407.
	1957	Distribution of the nearer bright stars in the colour-luminosity array, *Astron. J.*, **62,** 45.
	1958	The Hyades, Sirius, ζ Herculis, ϵ Indi, 61 Cygni and α Persei groups", *M.N.* **118,** 65 and 154, and *Observatory*, **78.**
Greenstein, J. L.	1956	The abundances of the chemical elements and the theory of their origin, *Publ. Astron. Soc. Pacific*, **68,** 185.
	1956	Evolution and white dwarfs, **68,** 501.
Harris, D. L.	1955	Photometry of the Lacerta Aggregate, *Ap. J.*, **121,** 554.
	1956	Colour-magnitude array of the white dwarfs, *Ap. J.*, **124,** 665.
Harris, D. L., Morgan, W. W. and Roman, N. G.	1954	Photometry and spectroscopy of stars in IC 348, *Ap. J.*, **119,** 622.
Haselgrove, C. B. and Hoyle, F.	1956	Age of type II stars, *M.N.*, **116,** 527.
Henyey, L. G., LeLevier, R. and Levée, R. D.	1955	Early stages of stellar evolution *Publ. Astron. Soc. Pacific*, **67,** 154.
Herbig, G. H. (Editor)	1957	*IAU Symposium on Non-Stable Stars*, Cambridge.
Hogg, A. R.	1953	Photometry of NGC 6025, *M.N.*, **113,** 746.
Hogg, A. R. and Kron, G. E.	1955	The galactic cluster IC 4665, *Astron. J.*, **60,** 365.
Hoyle, F.	1953	The fragmentation of gas clouds into galaxies and stars, *Ap. J.*, **118,** 513.
Huang, Su-Shu	1957	A dynamical problem and its bearing on stellar evolution, *Astron. J.*, **62,** 49.

JOHNSON, H. L.	1953	Magnitudes and colours in M 39 and NGC 752 and of selected visual binaries, *Ap. J.*, **117**, 353, 356, 361.
	1954	Magnitudes and colours in and near M 34 and IC 4665, *Ap. J.*, **119**, 181, 185.
	1954	Galactic clusters and stellar evolution, *Ap. J.*, **120**, 325.
	1957	Three-colour photometry of nearby stars, photometric distances of galactic clusters, and the colour-magnitude diagram for I Orionis, *Ap. J.*, **126**, 113, 121, 134.
JOHNSON, H. L. and SANDAGE, A. R.	1955	M 67 and its significance for stellar evolution, *Ap. J.*, **121**, 616.
	1956	Three-colour photometry in M 3, *Ap. J.*, **124**, 379.
JOHNSON, H. L., SANDAGE, A. R. and WAHLQUIST, H.	1956	The galactic cluster M 11, *Ap. J.*, **124**, 81.
JOHNSON, H. L. and HILTNER, W. A.	1956	Observational confirmation of a theory of stellar evolution, *Ap. J.*, **123**, 267.
JOHNSON, H. L. and MORGAN, W. W.	1953	Stellar photometry for standards of spectral type, *Ap. J.*, **117**, 313.
	1954	Heavily absorbed O-Association in Cygnus, *Ap. J.*, **119**, 344.
	1951	Colour-magnitude diagram of the Pleiades, *Ap. J.* **114**, 522.
JOHNSON, H. L. and SCHWARZSCHILD, M.	1951	Colour-magnitude diagram for M 15, *Ap. J.*, **113**, 630.
KRON, G. E., WHITE, H. S., and GASCOIGNE, S. C.	1957	Red and infrared magnitudes for 282 parallax stars, *Astron. J.*, **62**, 205.
KUKARKIN, B. V.	1954	Formation of stars, *Astron. Circ. U.S.S.R. No. 152.*
KUSHWAHA, R. S.	1957	Evolution of early main-sequence stars, *Ap. J.*, **125**, 242.
MASSEVICH, A. G.	1954	On the evolution of the stars: a summary of several papers in the *Astron J.* of the U.S.S.R., published in English by the Academy of Sciences of the U.S.S.R.
MESTEL, L. and SPITZER, L. JR.	1956	Star formation in magnetic dust-clouds, *M.N.* **116**, 503.
MCCREA, W. H.	1957	The formation of Population I stars, *M.N.*, **117** 562.
MORGAN, W. W., HARRIS, D. L., and JOHNSON, H. L.	1953	Characteristics of colour systems, *Ap. J.*, **118**, 92.
MITCHELL, R. I. and JOHNSON, H. L.	1957	Colour-magnitude diagram of the Pleiades, *Ap. J.*, **125**, 414.
MÜNCH, G.	1956	The age of early B-type supergiants, *Publ. Astron. Soc. Pacific*, **68**, 351.
NIKONOV, V. B., NEKRASOVA, S. V., POLOSUHINA, N. S., RACHKOVSKY, D. N. and CHUVAJEV, K. K.	1957	Colour-luminosity diagram for stars near the Sun, *Publ. Crimea Obs.* **17**, p. 42.
PARENAGO, P. P.	1954	Stars in the region of the Orion Nebula, *Publ. Sternberg Inst. Moscow*, **25.**
REDDISH, V. C.	1955	The masses of Population II stars, *M.N.* **115**, 32.
ROMAN, N. G.	1950	Dynamical characteristics of late F- and early G-type stars, *Ap. J.* **112**, 554.
	1955	High-velocity stars, *Ap. J. Supp.* Series I, No. 18.
ROMAN, N. G. and MORGAN, W. W.	1950	The moving cluster in Perseus, *Ap. J.*, **111**, 426.
SALPETER, E. E.	1955	Luminosity function and stellar evolution, *Ap. J.*, **121**, 161.
SANDAGE, A. R.	1954	Les processus nucléaires dans les astres, Liège.
	1953	Colour-magnitude diagram for M 3, *Astron. J.*, **58**, 61.
	1956	Luminosity functions, *Ap. J.*, **125**, 422, 435.
	1957	Stellar evolution, *Ap. J.*, **126**, 355.
SANDAGE, A. R. and WALKER, M.	1955	Globular cluster NGC 4147, *Astron. J.*, **60**, 230.
SAVEDOFF, M. P.	1956	Observational consequences of O-star formation, *Ap. J.*, **124**, 533.
SCHWARZSCHILD, M. and SPITZER, L. JR.	1953	The evolution of stars and chemical elements in early phases of a galaxy, *Observatory*, **73**, 77.
SCHWARZSCHILD, M., SCHWARZCHILD, B., SEARLE, L. and MELTZER, A.	1957	Spectroscopic comparison of high- and low-velocity K-giants, *Ap. J.*, **125**, 123.

SHARPLESS, S.	1954	Study of the Orion aggregate of early-type stars, *Ap. J.*, **119**, 200.
STOY, R. H.	1953	Magnitudes and colours of southern stars, Monthly Notices of the Astron. Soc. of South Africa. (And later.)
STRÖMGREN, B.	1952	Stellar evolution. *Astron. J.*, **57**, 65.
STRÖMGREN, B. and GYLDENKERNE, K.	1955	Spectral classification of G- and K-stars through photometry, *Ap. J.*, **121**, 43.
STRUVE, O.	1948	*Stellar evolution.* Princeton University Press.
TAYLER, R. J.	1954	Evolution of a massive star, *Ap. J.*, **120**, 332.
	1956	Evolution of unmixed stars, *M.N.*, **116**, 25
VYSSOTSKY, A. N. and DYER, E. R.	1957	Population differences among M-dwarfs, *Ap. J.*, **125**, 297.
WALKER, M. F.	1956	The extremely young cluster NGC 2264, *Ap. J.*, *Supp.* Series 2, No. 23.
	1957	The extremely young cluster NGC 6530, *Ap. J.*, **125**, 636.
WOOLLEY, R.v.d.R.	1958	Motions of the nearest stars *M.N.* **118**, 45.
WOOLLEY, R.v.d.R., and EGGEN, O. J.	1958	Colour-luminosity arrays of nearby stars; the Lacerta Aggregate, *Observatory*, **78**, 149.

The Visual Milky Way

First visual representation of the whole Milky Way, with bright stars and Magellanic Clouds; the Omega Centauri Cluster; a rare aurora.

Sergei Gaposchkin
Harvard College Observatory, Cambridge, Mass., U.S.A.

I. Introduction

If the astronomers in the last century had possessed, in the world of stars, an artistic genius equal to their intellectual grandeur, the realization of the fact that the dramatic multitude of stars is located not in the vicinity of the Sun, but in the direction of Sagittarius-Scorpius, would have dawned not in the twentieth but in the nineteenth century. A visual representation of the whole Milky Way would have revealed this at once.

II. Observations of the Milky Way

While I was in Australia and lived on the mountains near Canberra, I could observe the Milky Way during the seasons as it steadily covered larger and larger portions of the sky. Already on the way to Australia I had noticed, while in the southern Pacific Ocean, a special pattern of the regions of Ophiuchus, Sagittarius, Scorpius, Centaurus, Crux and Carina; a pattern which I could never observe either directly from the Northern Hemisphere, or vicariously reconstruct from the photographs. The facts that the centre of the Galaxy culminates in Australia (at Mount Stromlo) practically overhead, and that the unobscured sky was often transparent down to the very horizon, gave favourable conditions for observing the greater part of the Milky Way from one place. The major body of luminous and dark matter stretching from about α Aurigae over the centre to beyond α Lyrae, has been thus inspected from Mt. Stromlo in the period from October 1956 till May 1957.

I have first plotted the bright stars of the Milky Way, and then I have drawn with a blue pencil its dark regions, and with a red pencil its bright regions. Usually one could see a large portion of the Milky Way, and so I was able to compare intensities of different and far separated regions situated at the same altitude. As the season changed, different regions of the Milky Way came into sight; and I was able to interlock these regions in order to obtain more or less the same magnitude-scale all along the Southern Milky Way. During the second half of 1957, I repeated the same procedure in respect to the rest of the Milky Way in the constellations of Perseus, Cassiopeia, and Cygnus.

Since my primary programme in Australia consisted of obtaining spectra of emission stars, I combined my spectroscopic work with the drawing of the Milky Way in the following manner. During the exposure I worked in complete darkness

and, having left the star's image to drift along the slit, inspected the Milky Way through the opening of the dome, making corresponding drawings. (In drawing or perceiving the fine filament of the Milky Way, it is essential that the eye should be exposed to the darkness for at least 15 minutes, and that the angle of vision should be not less than 60°). After sketches of a few regions of the Milky Way had been finished in red and blue, I re-drew these regions on a sheet of paper of some five feet in length with pen and indian ink. The density of identical strokes was made proportional to the observed intensities of the bright areas. Thus the original drawing gives black regions as white; the negative of this drawing represents the Milky Way as it is actually seen.

The brightness of the regions was quantitatively expressed in terms of that of the Magellanic Clouds. The visual apparent magnitude of the Large Magellanic Cloud was assumed by me to be $-1^{m}0$ and that of the Small Magellanic Cloud $+1^{m}5$, figures previously determined by me; (see also Fig. 3). For example, the brightest area of the Milky Way lies around η Carinae and the next brightest is in Sagittarius. The former is $-3^{m}5$, the latter $-3^{m}2$.

III. Properties of the Milky Way in the Visual Light

A simple "staring" at the Milky Way (to use the phrase of Dr. Robert Oppenheimer) reveals many astounding features, some of which have never been recorded.* All of them are presented in Fig. 1.

1. Contrary to expectations, the detailed inspection of the dark matter reveals an astonishing regularity in its distribution. The dark matter appears to form a continuous belt, whose length and breadth is broken and twisted in a few places.

2. The dark belt pierces the Milky Way in its geometric centre and thus divides it into two almost equal parts; the northern one with its brightest region in Carina, and the southern one with its corresponding region in Sagittarius.

3. The dark belt in the constellation of Vela and Carina shows a strange form resembling that of an "S". It bridges both halves of the Milky Way from ϵ Carina in the south to λ Velorum in the north.

4. The well-known "Coal Sack" is not an isolated formation nor is it alone. Its neck is clearly connected with another similar formation, the "Coal Tin" in Carina, which in turn flows into the lower part of the "S" formation. The base of the "Coal Sack" is connected with the "Ditch", the huge dark masses to the left of α Centauri.

5. The "Ditch" begins at α Centauri, runs first smoothly along the middle of the Milky Way, and then, before the Norma region, curves upwards and reaches α Scorpii.

6. Corresponding to the "Ditch," which is on the right side of the centre of the Milky Way, is a similar "Sweep" of dark matter on the left side of it. This "Sweep", broader than the "Ditch", begins at about Nova Aquilae 1918, and ends at the

* Extensive visual-photometric work on the appearance of the Milky Way was carried out some 30–40 years ago by A. Pannekoek ("Die nördliche Milchstrasse" in *Ann. van de Sterrewacht te Leiden*, vol. II, 1920; and "Die südliche Milchstrasse" in *Ann. van de Boscha-Sterrewacht, Lembang-Java*, vol. 2, 1. Gedeelte, 1929). It is interesting to compare his drawings, which are essentially concerned with six different Milky Way regions, with the results illustrated in the present paper.

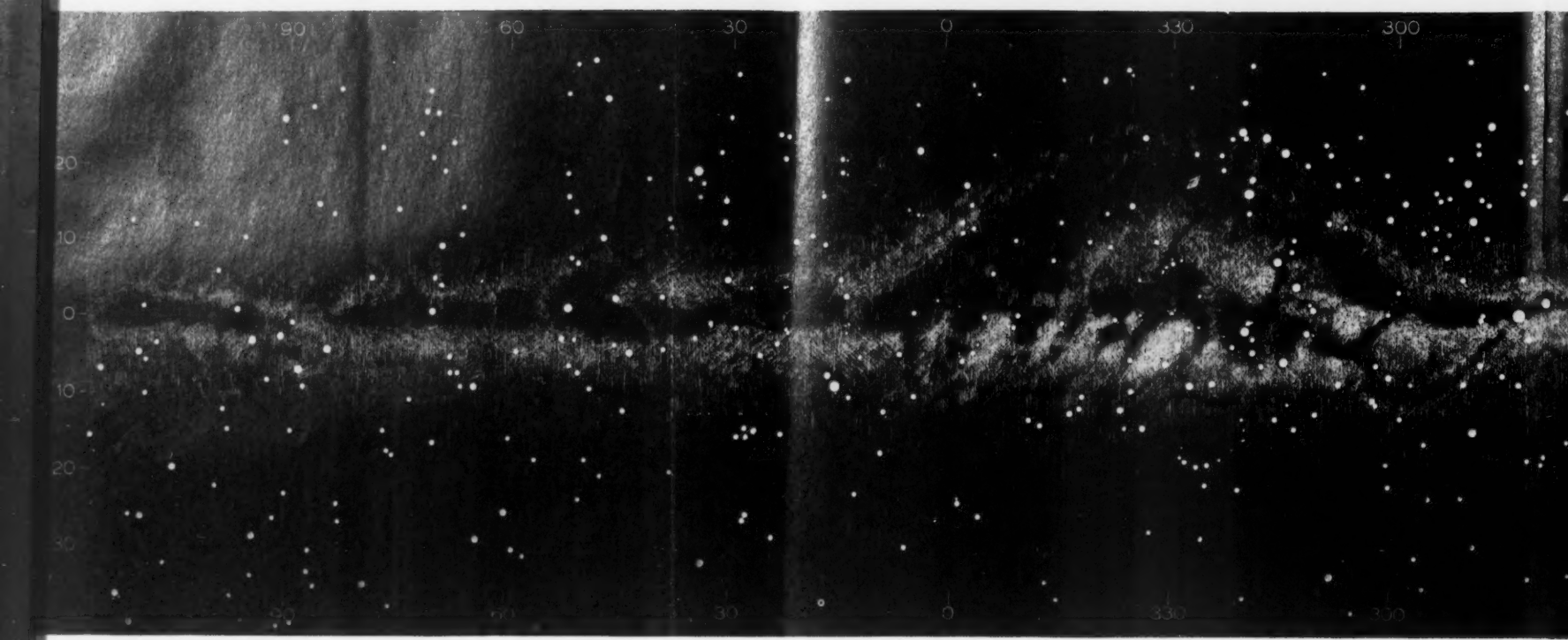
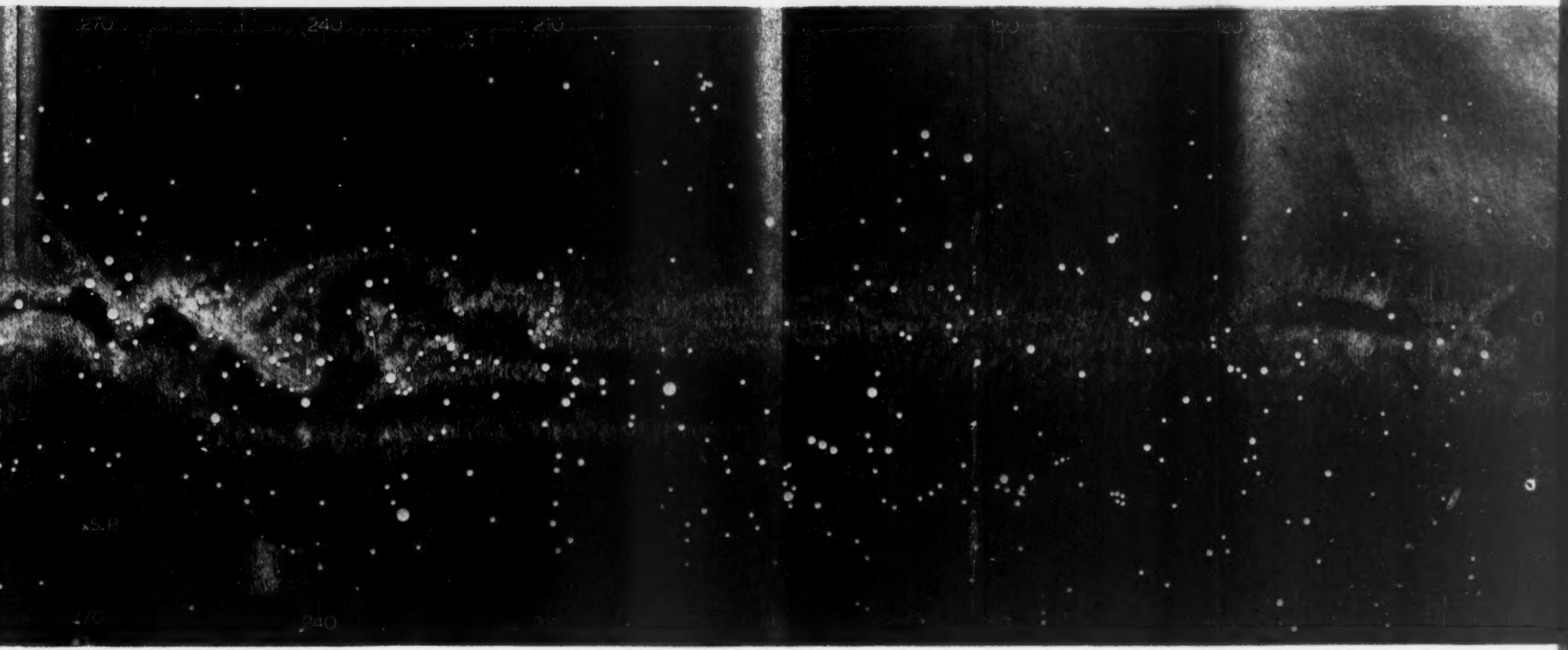


FIG. 1. The first visual representation of the whole
Clouds; (faithfully drawn by SERC

Symbols: ▲ = ω Centauri and 47 Tucanae; ↓ = v
in Ophiuchus, and the Andromeda Nebul
horizontally, latitudes vertically.



whole Milky Way, with bright stars and Magellanic
y SERGEI GAPOSCHKIN in 1956–57).
= variable stars; × SP = South Pole; Saturn
Nebula are drawn; galactic longitudes are plotted

[facing page 20

VOL
3
196

region close to α Scorpii (in our drawing this is occupied by the Planet Saturn) where the "Ditch" and the "Sweep" coalesce.

7. The "Ditch" and the "Sweep" outline perfectly the central part of the Milky Way —the "Bulge"—which is seen as the shimmering arch about the belt of the dark matter. There is an additional "ditch", thin and faint, to the South of α Centauri; there is no such "ditch" on the left side of the centre of the Milky Way.

8. The very middle of this "Bulge" is irregularly obscured by six "Lakes" of dark matter, which on the left side easily flow into the "Sweep" of Ophiuchus, but which on the right side are jammed by the narrow straits at γ_1 and γ_2 Normae, after which they join the "Ditch".

9. The separation of the luminous body of the Milky Way is clear throughout with one exception in the region of Auriga and Perseus. In this region the "Shine" of the Milky Way is the weakest. This is the "Tail" of the arm which is seen undivided in Gemini and Auriga—the "Backyard" of the Sun. The position in this region of the supergiant eclipsing system ϵ Aurigae, which is at least 3,000 parsecs from the Sun, indicates that the second lane outside is thin indeed.

10. This luminous "Backyard" joins on the right with many "holes" in Perseus, Cassiopeia, Cygnus, Aquila, and Lyra, and on the left with the undefined region of the "Dogs" and Monoceros. The latter transform themselves into the most distorted regions of the Milky Way, those of Puppis, Vela, and Carina.

11. There is no gainsaying that there is only one region—that around Sagittarius—which may be regarded as the central part of the Milky Way. However, the dramatic regions to the right of the centre, especially that in the region of the "S" formation, are difficult to visualize as a superstructure of lanes: the luminous part is too wide as compared with that on the left side. It is as if a foreign body enters the Milky Way here.

12. The brightest part of the bulge is just below the largest dark lake in Ophiuchus (close to θ Ophiuchi). Unfortunately the very centre of the Milky Way is washed over by dark lakes.

13. Towards the South there is no luminous formation analogous to that towards the North. This absence indicates again that the Sun is above the plane of the Milky Way. (The first indication was that the Northern luminous part was definitely wider than that of the South.)

14. There are quite a few "sweeps" and "arches", black and white, which are significant, notably one bright sweep from about η Carinae to u Velorum.

15. The most brilliant part of the Milky Way is around η Carinae and the next brightest is that in Sagittarius.

16. The Magellanic Clouds are, as seen perspectively, not far from the luminous haze of the Milky Way. This is especially true for the Large Magellanic Cloud. It has a distinct spiral or "S" shape, with the major axis directed perpendicularly toward the Milky Way, pointing to the star γ Volantis. The Large Magellanic Cloud has a definite dark lane towards us and is tilted by 55°.5 (90° is *en face*). The Small Magellanic Cloud has no regularity of form and looks like a distorted "Knackwurst".

17. The possible connections between the Large Magellanic Cloud and the Milky Way are: (a) from β Carina and over γ Volantis; (b) from γ Velorum and under Canopus to δ Doradus; (c) from α Trianguli Australis over the South Pole.

18. The total apparent visual magnitude of the Milky Way is between $-5^{m}0$ and $-6^{m}0$. That is why, when the regions of Sagittarius, Centaurus, Crux, and Carina are high above in the moonless and cloudless night, the light is sufficient to illuminate your way over the stony ground from the telescope to your flat.

IV. The Visual Profile of the Milky Way

The graphical reconstruction of the Milky Way has previously been done either by analogy with other galaxies, or by interpretive explorations of the distant regions of the Milky Way, notably by radio. The first method is similar to an attempt to draw a plan of a city, say that of Washington, D.C., from studying the plan of another city, say Moscow, U.S.S.R. The second method recalls the situation of an expedition which attempts to explore the distant frontiers of the ocean before investigating the harbour of departure.

The visual profile of the Milky Way can easily be reconstructed from a schematic picture representing the main features of the system. These, as we have indicated above, are: Holes, Sweep, Bulge, Lakes, Ditch, Coal Sack, Coal Tin, "S"-shape, Backyard, and the Tail.

The circular arrangements of bright and dark regions around the Sun represent, after some modifications, the actual profile of the Milky Way as seen from a point in space far beyond the Sun. Indeed, the central part of the Milky Way will appear from that distant point as it appears from the Sun. The thin and weak shimmer of the "Backyard", superimposed on the central part, will produce only small changes: there will be a thin veil over the regions of "Lakes", so that the upper bright areas will becomes slightly brighter and the dark lakes will be transformed into a continuous path of dark matter. The "Sweep" in Ophiuchus and the "Ditch" in Centaurus will suffer some changes; their lower parts will be covered with a bright haze and their top parts will practically disappear. The contrast between the "Coal Sack" and the "Coal Tin" appears less; the "S"-shape will be cut in two by the bright regions from the "Dogs" to Puppis, and the "Hole" in Cygnus will be still darker, being covered by the "Hole" in Cassiopeia. The central "Bulge", with its shimmer, will remain entirely unaffected in the North, but its shimmering counterpart must appear in the South because the "Bulge" is symmetrical in structure. Some of the brightest stars, of about $-7^{m}0$, will be seen along the central dark band of the Milky Way. I have indicated a few of them in the second part of Fig. 2. It should be particularly noted that ϵ Aurigae, which is found originally in the "Backyard", will now be projected to a position in front and slightly to the left of the "Bulge".

V. Additional Visual Observations

During the same observing period visual impressions were recorded of the Large Magellanic Cloud (reproduced in Fig. 3), the globular cluster ω Centauri (Fig. 4) and a rare auroral phenomenon (Fig. 5).

VOL.
3
196

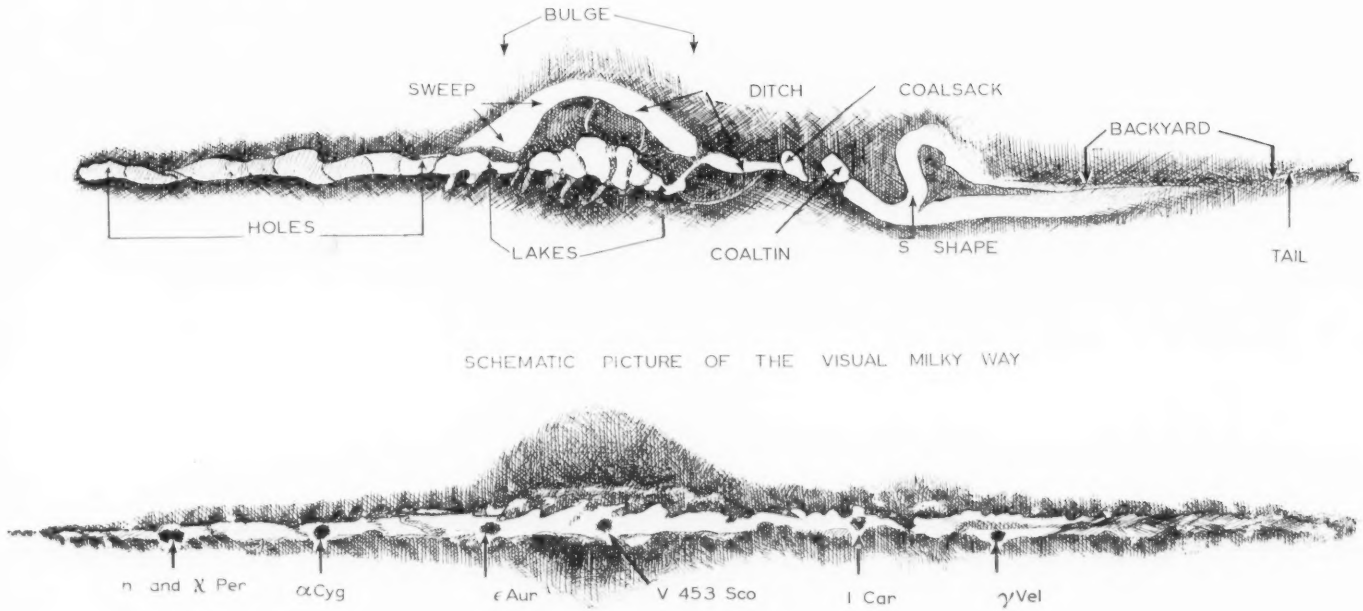


FIG. 2. The visual profile of the Milky Way. (SERGEI GAPOSCHKIN, 1956–57.)

FIG. 3. Visual representation of the Large Magellanic Cloud.
From observations on several nights of favourable conditions, I have made this drawing of the Large Magellanic Cloud. Unmistakably it is the shape of an "S", with small protruding filaments which might connect it to the Milky Way.

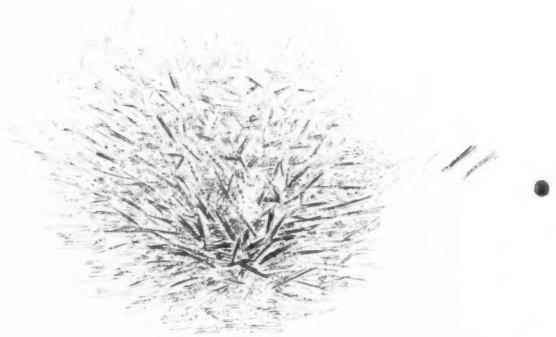


FIG. 4. Visual impression of the globular cluster ω Centauri.
Several times I was startled by the appearance of ω Centauri through a low-powered telescope. Many of the stars seemed to form small streaks of light, and the centre revealed a dark area, void of stars. This suggests that it is an old cluster in the process of disruption. The globular cluster 47 Tucanae appeared in the usual form and not at all similar to ω Centauri.

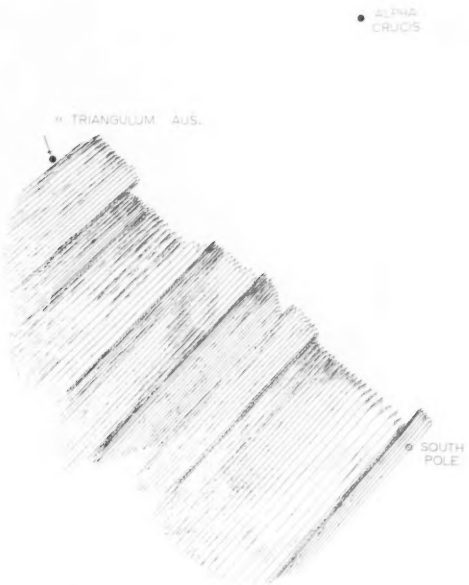


FIG. 5. A rare aurora observed on March 10–11, 1957.
On that night, at 3.00 a.m., there was a display of Southern Aurora, shaped in a way I have never observed in the United States, and have not seen represented in photographs. It appeared as a collection of picket fences, all slanted at an angle of about 45° towards the horizon. Each picket was definitely marked in shape. The colour was usually weak: faint orange-brown in the joined areas, and whitish-gray in the rest. Within fifteen minutes the whole phenomenon had undergone a complete change, finally disappearing.

Star Formation and the Galactic Magnetic Field

L. Mestel

St. John's College, Cambridge

Summary

A strong galactic magnetic field, as postulated in much current astronomical theory, must play an important rôle in the condensation of interstellar matter. In this article we consider the effect of introducing a magnetic field into three problems that have been much discussed in the literature: the condition of gravitational binding; the rate of accretion by a condensation already formed; and the evolution of a rotating disc. In all three cases the results are significantly altered by the presence of the field.

1. Introduction

There is now impressive evidence for a strong, large-scale galactic magnetic field, with an energy comparable with the kinetic energy of the mass motions observed, and probably greater than the thermal energy. The current density $\boldsymbol{j}$ maintaining a field $\boldsymbol{H}$ is $c\nabla_\wedge \boldsymbol{H}/4\pi$, and so the force density exerted by the field is $(\nabla_\wedge \boldsymbol{H})_\wedge \boldsymbol{H}/4\pi$; hence if the field has a sufficiently large scale, so that the currents flow in circuits of large dimensions, the magnetic forces could still be small even if the field strength is large. But in the problem of star formation, we are concerned with the compression of tenuous galactic gas into a much smaller volume. Because of the strong coupling between matter and field, condensation in general implies a strong local distortion of the field and the generation of strong forces.

Without asserting that earlier treatments of all problems in cosmical gas dynamics are obsolete, it is certainly imperative to reconsider results obtained by neglecting all forces except gravitation, gas pressure and centrifugal force. The introduction of magnetic (and indeed centrifugal) force seriously complicates the treatment, and detailed results will require the use of automatic computers. For the moment, however, exploratory approximate discussion is worth while, if only to help us to decide what problems to programme for machine solution.

In this article we consider the effect of a magnetic field on three aspects of the star formation problem. First, the difficulty of forming small gravitationally bound masses from strongly magnetic matter is resolved by showing that matter and field do not move together in all circumstances. We then assume that a contracting proto-star has been formed, and discuss the effect of the field on the rate of gravitational accretion (formation of a massive star from a small one by accretion being regarded as just formation in two stages). Finally, we illustrate the probable importance of a magnetic field in removing excess angular momentum from a condensation by a rather idealized discussion of the evolution of a rotating disk with an internal field.

2. The Break-up of a Magnetic Cloud into Stars

When a strong galactic magnetic field was first postulated, one of the objections raised by theoretical astrophysicists was that the field made star formation virtually

VOL.
3
196

impossible. It was argued that the field is strongly coupled to the gas, so that if a globule of radius R_0, with a roughly uniform field H_0, contracts under gravity to a radius R, then the field increases to

$$H = H_0\left(\frac{R_0}{R}\right)^2. \qquad \ldots . (1)$$

The force density exerted by the field is $(\nabla \wedge \boldsymbol{H}) \wedge \boldsymbol{H}/4\pi$, which is of order $H^2/4\pi R \simeq (H_0 R_0^2)^2/4\pi R^5$. The thermal pressure gradient in a cloud of temperature T and mean molecular weight μ is $\mathcal{R}\rho T/\mu R$, $\mathcal{R}$ being the gas constant and ρ the density. While the cloud is diffuse the opacity will be low, and internal radiators—dust grains, molecules—will eliminate heat generated by compression, and keep the temperature nearly constant under contraction. Hence the pressure gradient increases like $1/R^4$, while the gravitational force density $\simeq 3GM^2/4\pi R^5$. In order that the globule may begin to contract, the gravitational force must exceed the opposing pressure gradient and magnetic force at the initial radius R_0. Under contraction the pressure becomes less and less important, but the gravitational and magnetic forces increase proportionately. A lower limit to the mass able to hold itself together is given by ignoring the pressure, and equating the magnetic and gravitational forces:

$$M = \frac{H^3}{(3G)^{\frac{3}{2}}}\left(\frac{3}{4\pi\rho}\right)^2, \qquad \ldots . (2)$$

where H and ρ are simultaneous values. Putting $H = 10^{-6}$ when $\rho = \frac{5}{3} \times 10^{-23}$ (a conservative estimate for the field), we find for (2) approximately $10^3 \odot$.

One way of getting over this difficulty is to postulate much more rapid contraction along the field than across it, so that high densities are built up without a simultaneous large increase in field strength. This could conceivably occur with the help of an external non-spherical pressure field, but it is difficult to see why such a condensation should occur spontaneously. Once the cloud has taken the shape of a rough spheroid, oblate to the direction of the field, the gravitational force along the field will be the same as the magnetically diluted force across it, and uniform contraction will ensue: excess contraction along the field would increase the density, and therefore also the net lateral force.

We are therefore forced to look for a means by which the strong coupling of the field and the matter may be relaxed. The theorem of the freezing of the field into a moving conducting gas results from combining the induction equation

$$\nabla \wedge \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t} \qquad \ldots . (3)$$

with the equation

$$\boldsymbol{E} + \frac{\boldsymbol{\upsilon} \wedge \boldsymbol{H}}{c} = 0, \qquad \ldots . (4)$$

where $\boldsymbol{E}$ is the electric field measured in a fixed frame of reference and $\boldsymbol{\upsilon}$ is the velocity of the gas. In a fully ionized gas the small terms dropped from (4) are the

Ohmic field, the Hall field, and the pressure gradient term; it is easily shown that they are negligible on a cosmical scale. Equations (3) and (4) yield

$$\frac{\partial \boldsymbol{H}}{\partial t} = \nabla \times (\boldsymbol{\upsilon} \wedge \boldsymbol{H}); \qquad \ldots . (5)$$

it is well known that (5) implies the conservation of flux in a circuit moving with the gas.

Equation (4) is essentially the equation of motion of the electron gas relative to the ions. In an H I cloud the plasma of ionized metals is only a small fraction of the total gas; the problem must therefore be considered anew, with the velocity of the plasma carefully distinguished from that of the gas as a whole. In a strongly magnetic, lightly ionized cloud, the equation to the plasma is approximately

$$\frac{\boldsymbol{j} \wedge \boldsymbol{H}}{c} + n_i \boldsymbol{F}_{iH} = 0, \qquad \ldots . (6)$$

where n_i is the ionic density, and $\boldsymbol{F}_{iH}$ the mean force on an ion due to collisions with the neutral hydrogen. The magnetic force causes just that drift of plasma relative to the neutral gas for friction and magnetic force to balance.

The equation to the motion of the ions is approximately

$$\boldsymbol{E} + \frac{\boldsymbol{\upsilon}_i \wedge \boldsymbol{H}}{c} + \frac{n_i \boldsymbol{F}_{iH}}{n_e e} = 0, \qquad \ldots . (7)$$

with n_e the electron density, $-e$ the electronic charge, and $\boldsymbol{\upsilon}_i$ the velocity of the ions. From (6) and (7) we have

$$\boldsymbol{E} + \frac{\boldsymbol{\upsilon}_i \wedge \boldsymbol{H}}{c} - \frac{\boldsymbol{j} \wedge \boldsymbol{H}}{n_e e c} = 0. \qquad \ldots . (8)$$

If the field is axially symmetric the poloidal component of $\boldsymbol{j} \wedge \boldsymbol{H}/n_e e c$ leads merely to a very slight departure from the law of isorotation, while a toroidal component interacts strongly with the rotation field in the cloud in such a way as to cancel itself out. In any case, even if there is a non-vanishing toroidal component, so that the outward motion of the field is not strictly identical with that of the plasma, the relative drift is of order $-j/n_e e = V$, the velocity of drift of the electrons relative to the ions. On using $j \simeq cH/4\pi R$, it is easy to verify that in a gas cloud $\tau = R/V$ is far too long for the drift of field relative to plasma to be significant, even if the plasma density is far below that found in normal H I clouds. Hence we conclude that (5) holds, provided we write $\boldsymbol{\upsilon}_i$ instead of $\boldsymbol{\upsilon}$: the magnetic field moves with the *plasma*, while the drift of plasma plus field through the neutral gas is given by (6).

In a gravitationally contracting H I cloud, the magnetic force will try and prevent the plasma from following the motion of the neutral gas, whereas the mutual friction tends to force the two gases to move together. Hence, if the frictional coupling is strong, the relative drift of plasma and neutral gas allowed by (6) is small compared with the freely falling cloud velocity, the two gases effectively move together, and the field is frozen into the cloud. On the other hand, if the coupling is weak, the neutral gas can flow across the field without dragging the plasma with it; hence the neutral gas will increase in density without a simultaneous large increase in the magnetic field strength.

If the number density of ions and electrons is approximately one ten-thousandth that of the neutral hydrogen, as in a normal H I cloud, it is found that the relative drift of plasma and neutral gas is only about one per cent of the free-fall velocity. We therefore need rather special conditions for uncoupling of the matter and the field. One possibility arises if the cloud is shielded from galactic radiation by a high density of absorbing dust grains; the plasma density then decays steadily through recombinations, or, more efficiently, through attachment to the dust grains. It is found that the time of decay of the plasma is less than the time of free-fall, and so we may expect frictional uncoupling to occur.

If, however, the plasma density remains too high, then a freely falling cloud will drag the field with it, and break-up into stars will be forbidden. Two other possible ways of losing excess magnetic energy suggest themselves:

(1) The angular momentum in the disk leads to the build-up of a centrifugal field which slows up the contraction, giving the plasma and field more time in which to slip out. The objection to this is that the field will probably be too efficient in removing angular momentum (cf. Section 3), so that if centrifugal force is initially negligible it will remain so.

(2) The cloud falls freely under gravity until the opacity becomes too high for the compressional heat generated to be radiated away. The contraction is then determined by the rate of flow of energy down the equilibrium temperature gradient. Again, this may yield a sufficiently long time-scale for the field to force the plasma out of the cloud, thus decreasing the field energy enough for star formation to occur.

A fuller account of this work is presented in MESTEL and SPITZER (1956).

3. MAGNETIC FIELDS AND GRAVITATIONAL ACCRETION

Let us now suppose that the problem of star formation from magnetic matter has been solved, in that excess magnetic energy has been lost, and masses of stellar order can hold themselves together. We imagine that a proto-star has formed, and is slowly contracting in the KELVIN-HELMHOLTZ time-scale. We have seen that uncoupling between the magnetic and matter fields demands rather special conditions; we may therefore suppose that the plasma density is high enough for the star to conserve its flux under contraction. Even if the proto-star is still shielded from galactic radiation, collisional ionization becomes important once temperatures are above $10^{3\circ}$ K or so, yielding a high plasma density and strong frictional coupling.

To simplify things, we assume a uniform field $\boldsymbol{H}_0$ far from the star, and a uniform field $\bar{\boldsymbol{H}}$ within the star parallel to $\boldsymbol{H}_0$; at the time of onset of freezing-in, $\bar{\boldsymbol{H}} = \boldsymbol{H}_0$, the stellar lines of force being just segments of the cloud field lines, which are themselves part of the local galactic field. As the star contracts from the radius R_0, $\bar{H}$ increases according to (1).

The gravitational field of the star causes inward flow of matter. In the absence of a magnetic field, the flow would be as described by BONDI for a star at rest—spherically symmetric accretion (BONDI, 1952). But as the field outside is distorted by the inflowing matter, it exerts a force on the plasma, and so indirectly on the neutral gas as well. A steady state is approached, in which the magnetic force just balances the component across the field of the non-magnetic forces acting on the plasma: a familiar situation in hydromagnetics. We assume that the steady-state field is purely poloidal. It can be shown that with the given boundary conditions at infinity and

at the stellar surface, a twisted field is probably unstable and would rapidly revert to a purely poloidal form; while apart from questions of stability, the transition from an initially poloidal field to a steady-state twisted field is forbidden by COWLING's dynamo theorem. Then the equation to the field is

$$\nabla \wedge \boldsymbol{H} = 4\pi \frac{\boldsymbol{F} \wedge \boldsymbol{H}}{H^2}, \qquad \ldots . (9)$$

where $\boldsymbol{F}$ is the density of the non-magnetic force acting on the plasma. The field is then given by the solution of (9) which reduces to $\boldsymbol{H}_0$ at infinity, and with a radial component continuous at the star's surface with $\overline{H}_r$.

The accurate solution of (9) is difficult because of the non-linearity. However, the approximate shape is as in Fig. 1. Far from the equator the field is approximately

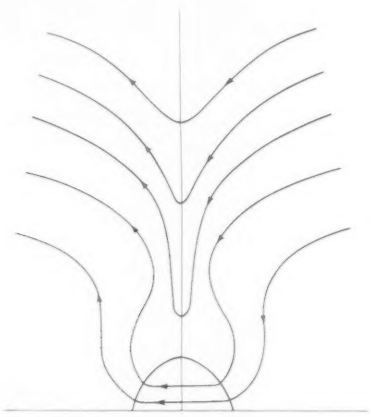


FIG. 1. Approximate structure of the magnetic field.

irrotational; the field is so strong that the currents flowing across it need only be very small for the magnetic force to balance the transverse component of $\boldsymbol{F}$. But the approximate solution $\nabla \wedge \boldsymbol{H} = 0$ cannot hold everywhere, as this would involve a neutral point at which the condition (9) could not hold. In fact near the equator the field is small but has a large curl, so that $(\nabla \wedge \boldsymbol{H}) \wedge \boldsymbol{H}/4\pi$ is of the same order as $\boldsymbol{F}$. An important point is that the stellar field lines remain part of the galactic field: there is no early breaking of some of the field-lines to form lines of finite length, as in the dipole field $\nabla \wedge \boldsymbol{H} = 0$, for this would necessarily involve a neutral point. Only when the curl of the equatorial field has become large enough to yield a significant drift of plasma across the field—so that the gravitational force is balanced by the plasma mass-acceleration—will a neutral point form. Even so, this change in the field's structure is unlikely to affect the rate of accretion.

If the cloud were completely ionized, the principal force $\boldsymbol{F}$ would be the gravitational pull of the star. Equation (9) then fixes the field, and the dynamics of the plasma flow along the field can then be studied. In a lightly ionized gas, however, the neutral matter is not tied to the field, but its drift across the field is resisted by the frictional drag of the plasma, which *is* tied to the field. If this friction is everywhere small compared with the gravitational force on the *neutral* gas, then the

neutral gas flows in approximately as in BONDI's theory, unimpeded by the field. The dominant force $\boldsymbol{F}$ on the plasma, determining the field by (9), will be the friction except for very low density clouds, when the friction will be less even than the gravitational force on the *plasma*. However, for clouds of moderately high density (greater than $\simeq 5\times10^3$ hydrogen atoms/cm^3 at infinity), the frictional drag on neutral matter flowing across the field with BONDI velocities is everywhere larger than gravity; hence for such densities the flow of the neutral matter across the field is much below BONDI's value, while the component along the field remains comparable with BONDI's value. Hence it is a good approximation to consider both gases as being forced to flow down the field lines; this will be assumed from now on. The principal force $\boldsymbol{F}$ acting on the plasma will again be the friction; but as the neutral matter flows across the field at just that rate which makes the friction balance the gravitational force on the neutral gas, the field will be given by (9), with $\boldsymbol{F}$ equal to the gravitational force on the whole gas. Thus for an H I cloud of high enough density, both the field and the flow of gas are given by equations identical with those for a fully ionized gas.

If a line of force intersects the star's surface, both gases flow straight into the star. Of those lines which cross the equator outside the star, consider first those which cross the equator beyond BONDI's critical radius $r_c = GM/c_\infty^2$ (c_∞ being the velocity of sound at infinity, which in the present case of isothermal flow is constant). A slight drift of gas along such a line towards the equator will establish hydrostatic equilibrium, and so no continuous flow of gas down the field will ensue; there will be just a slight drift of neutral matter across the field. On the other hand, for lines cutting the equator at $r \ll r_c$, the scale height for equilibrium is small compared with r_c. Only impossibly large densities at the equator would support the gas in equilibrium along the field line, and so we expect a continuous flow of both gases down the field towards the equator. A dense equatorial zone is built up, in which the plasma is neutralized by recombinations of ions and electrons. The neutral hydrogen and the neutralized plasma then flow in towards the star. Hence all the matter flowing down the field lines, whether directly into the star or onto the equator, ultimately reaches the star. The equation to the field within the zone is again (9), but with a much larger force density $\boldsymbol{F}$.

We now discuss the flow down the field lines cutting the equator within r_c. As the flow is everywhere parallel to the field we write

$$\boldsymbol{v} = \kappa \boldsymbol{H} \qquad \ldots\ldots (10)$$

where κ is a scalar function of position. In a steady state,

$$0 = \nabla.(\rho \boldsymbol{v}) = \nabla.(\rho\kappa \boldsymbol{H}) = \boldsymbol{H} \,.\, \nabla(\rho\kappa), \qquad \ldots\ldots (11)$$

so that

$$\rho\kappa = \text{constant along a line of force} = \rho_\infty \kappa_\infty. \qquad \ldots\ldots (12)$$

Hence in a steady state there is a non-vanishing velocity field $v_\infty = \kappa_\infty H_0$ far from the star; even if the gas is initially at rest at infinity, the flow will asymptotically approach a state with v_∞ finite. This is because the magnetic field prevents flow of gas from one flux-tube to another; hence if the field does not vanish at infinity, each flux-tube remains of finite width, with continuity demanding a finite velocity everywhere.

The equation to the steady flow along a field line is just BERNOULLI's equation with the gravitational field as a body force:

$$c_\infty^2 \log \frac{\rho}{\rho_\infty} - \frac{GM}{r} + \tfrac{1}{2} v^2 = \tfrac{1}{2} v_\infty^2. \qquad \ldots . (13)$$

In a steady state the magnetic field does no work, as the flow is parallel to the field. Introducing non-dimensional variables x, z, y, h, defined by

$$r = \left(\frac{GM}{c_\infty^2}\right) x, \quad \rho = \rho_\infty z, \quad v = c_\infty y, \quad H = H_0 h, \qquad \ldots . (14)$$

we have from (12) and (13)

$$\frac{yz}{h} = y_\infty, \qquad \ldots . (15)$$

and

$$-\frac{1}{x} + \log z + \tfrac{1}{2} y^2 = \tfrac{1}{2} y_\infty^2. \qquad \ldots . (16)$$

Eliminating z, we get

$$\phi(y) - \psi(x) = \gamma, \qquad \ldots . (17)$$

where

$$\left.\begin{aligned} \phi(y) &= \tfrac{1}{2} y^2 - \log y, \\ \psi(x) &= \frac{1}{x} - \log h, \\ \gamma &= \tfrac{1}{2} y_\infty^2 - \log y_\infty. \end{aligned}\right\} \qquad \ldots . (18)$$

Equations (15) and (17), applying to each line of force, replace BONDI's equations

$$x^2 y z = \lambda, \qquad \ldots . (19)$$

$$(\tfrac{1}{2} y^2 - \log y) - \left(\frac{1}{x} + 2 \log x\right) = -\log \lambda. \qquad \ldots . (20)$$

for spherically symmetric flow.

The discussion of (17) follows closely BONDI's discussion of (20). The minimum value of $\phi(y)$ is 0·5, and it is attained for $y = 1$; hence the parameter γ must be chosen so that for all x along the line of force considered

$$\gamma + \psi(x) \geqslant 0{\cdot}5. \qquad \ldots . (21)$$

Now $\psi(x) \to 0$ as $x \to \infty$; also, for all physically relevant fields, $\psi(x) \to \infty$ as $x \to 0$. Suppose first that $\psi > 0$ for all finite x. Then (21) is satisfied for $\gamma \geqslant 0{\cdot}5$, and $\phi(y) \geqslant 0{\cdot}5$. By inspection of the graph of $\phi(y)$ against y, we have

$$\begin{aligned} &\text{if } y_\infty > 1, \text{ then } y > 1 \text{ for all finite } x; \\ &\text{if } y_\infty < 1, \text{ then } y < 1 \text{ for all finite } x; \\ &\text{and if } y_\infty = 1, \begin{cases} \text{either } y > 1 \text{ for all finite } x; \\ \text{or} \quad\ \ y < 1 \text{ for all finite } x. \end{cases} \end{aligned}$$

If ψ is non-negative for all x, but has a zero for $x = x'$, then the solution with $y_\infty = 1$ will by (17) attain the sonic velocity again at x'; hence a solution supersonic beyond the zero may change over to a subsonic solution at x', and *vice versa*.

Now suppose ψ becomes negative for some range of x along the line of force considered, and let x_m be the point at which it attains its minimum. Then the inequality (21) will hold everywhere along the line provided it holds at x_m. Thus, if $\gamma+\psi(x_m) = 0{\cdot}5$, then $y(x_m) = 1$, and the flow on either side of x_m could be supersonic or subsonic. The velocity at infinity is then given by

$$\phi(y_\infty) = \gamma = (0{\cdot}5 - \psi(x_m)); \qquad \ldots . (22)$$

approximately,

$$y_\infty = (1 - 2\psi(x_m))^{\frac{1}{2}} \text{ if the flow is supersonic at infinity;} \qquad \ldots . (23)$$

or

$$y_\infty = \exp(\psi(x_m) - 0{\cdot}5) \text{ if the flow is subsonic at infinity.} \qquad \ldots . (24)$$

All values of y_∞ yielding $\phi(y_\infty)$ larger than (22) are allowed, as then $\gamma+\psi(x_m) > 0{\cdot}5$, but values of y_∞ between (23) and (24) will not yield steady state solutions. If y_∞ is less then (24), then the flow is subsonic everywhere; if greater than (23) it is supersonic everywhere. For either of the two values (23) and (24) the flow is just sonic at x_m, and so flow supersonic for $x > x_m$ may become subsonic for $x < x_m$, or by a discontinuity in the velocity gradient at x_m, may remain supersonic for $x < x_m$. Similarly, flow subsonic for $x > x_m$ may become supersonic or remain subsonic for $x < x_m$.

However, not all steady state solutions are physically plausible. As $\psi(x)$ gets large for small x, by (17) the velocity in the solutions subsonic near the star, becomes very small, and leads by (15) to very large densities. This is again due to the smallness of the scale-height within the BONDI sphere: for low velocity solutions, the equation of motion approximates to the equation of hydrostatic support. If such a steady state, with a large outward pressure gradient, existed initially, then it might be stable and persist, the pressure gradient absorbing the gravitational potential energy released as matter slowly drifts in along the field line. But if initially the pressure gradient is small compared with the gravitational force density, then matter within the Bondi sphere will fall freely along the field line towards the star, and only solutions supersonic near the star will be relevant.

Assuming, then, that the flow near the star is supersonic, the problem arises as to which of the allowed steady state solutions will arise in practice. Let us suppose that initially the cloud is at rest at infinity. Within the BONDI sphere the gravitational field causes inward flow, and the region of flow is gradually extended outwards by continuity. Once the system has reached that velocity field, supersonic near the star, consistent with the equations of motion and continuity, and with the *least energy per gram of matter*, there is no reason why the velocities should increase above this level. Hence we expect the system to approach the flow with the minimum allowed velocities: if $\psi(x) \geqslant 0$ everywhere, we take $y_\infty = 1$, while if $\psi(x)$ has a negative minimum at x_m, we take $y(x_m) = 1$, and adopt the solution that is subsonic for $x > x_m$.

However, it should be emphasized that the equations of motion and continuity do not forbid all higher velocities at infinity, with correspondingly greater flow of matter down the field lines. If $\psi(x) \geqslant 0$, all supersonic velocities are allowed; while

if $\psi(x) < 0$ for some range of x, all values of y_∞ greater than (23) are allowed. Solutions of this type, with y_∞ greater than the minimum allowed, will be relevant if the cloud is not initially at rest at infinity, but has matter being driven in rapidly enough by an external pressure field (e.g. by an external H II region). This is in striking contrast to BONDI's spherically symmetric flow, and is again due to the confining effect of the magnetic field. If BONDI's cloud had a non-vanishing inward velocity at infinity, the radial flow would rapidly build up a high pressure gradient and dam back the flow, which would asymptotically approach the steady pattern with zero velocity at infinity. In the present problem, the magnetic field prevents the build-up of such a density gradient, by restricting the flow to be along the field. The flow along each infinitesimal flux-tube is analogous to flow through a pipe of variable but finite cross-section: not only are there low velocity solutions, in which the pressure controls the flow until it becomes sonic, but there are also high velocity solutions in which the pressure gradient is negligible everywhere. These results are analogues of well-known results in laboratory gas dynamics.

In BONDI's spherically symmetric problem there is a definite upper limit to the total rate of accretion by the star:

$$A \simeq 4\pi\left(\frac{GM}{c_\infty^2}\right)^2 c_\infty \rho_\infty = 4\pi r_c^2 c_\infty \rho_\infty \qquad \ldots . (26)$$

Again, a higher rate of inflow would lead to a large pressure gradient beyond the sphere r_c, which would dam back the flow. In the present case, the rate of inflow across a sphere surrounding the star, with outward unit normal $\boldsymbol{n}$, and of radius less than r_c, is

$$-\int_S \rho \boldsymbol{v} \,.\, \boldsymbol{n} dS = 2\int_{S'} |\kappa| \rho \boldsymbol{H}.\boldsymbol{n}\, dS = 2\rho_\infty \int_{S'} \frac{v_\infty}{H_0}(\boldsymbol{H}.\boldsymbol{n}) dS, \qquad \ldots . (26)$$

where S' is the northern hemisphere. We have seen that there exist steady state solutions with v_∞ supersonic; these yield a rate of flow across S greater than

$$\frac{2\rho_\infty c_\infty}{H_0}\int_{S'} \boldsymbol{H}.\boldsymbol{n}\, dS = 2\pi L_\infty^2 c_\infty \rho_\infty, \qquad \ldots . (27)$$

where L_∞ is the radius at infinity of the flux-tube containing S. The total accretion rate is given when S is just the BONDI sphere; hence if L_∞ for this sphere exceeds $\sqrt{2}r_c$, the accretion rate exceeds BONDI's maximum (25). Even if L_∞ were less than $\sqrt{2}r_c$, high enough velocities at infinity would still lead to enhanced accretion.

However, supersonic velocities at infinity will occur only if there is an external pressure field driving the matter in. We have seen that a cloud initially at rest at infinity will spontaneously approach a state with at most the sonic velocity at infinity. If the freezing-in of the field into the proto-star starts at a radius $R_0 > \sqrt{2}r_c$, then $L_\infty > R_0 > \sqrt{2}r_c$, and so the "cross-section" for capture of material will be greater than the area of BONDI's sphere; but this will not lead to an enhanced accretion rate unless the velocities at infinity remain of sonic order. This depends on the behaviour of the function $\psi(x)$ along the line of force considered. If ψ remains positive, then $y_\infty = 1$; but if the field strength increases too rapidly inwards along

the line of force (implying a sharp decrease in the area of the flux-tube along which the gas flows), ψ becomes negative, and y_∞ is by (24) strongly subsonic.

We conclude that a magnetic field can assist accretion of matter by a proto-star; but we cannot decide whether the system will *spontaneously* accrete more than BONDI's maximum without a detailed study of the structure of the magnetic field in the cloud. However, approximate estimates suggest that the narrow flux-tubes near the equator, that "pinch" the equatorial plasma, will by (24) strongly cut down the velocity at infinity, so that the spontaneous accretion-rate is likely to be well below BONDI's maximum.

4. MAGNETIC FIELDS AND STELLAR ROTATION

If rotation could be neglected, the theory of Section 2 could be applied at all stages of the star's contraction to the main sequence, provided the correct magnetic field, fitting on to the changing field within the proto-star, is used at each epoch. However, all clouds have some angular momentum, and so we may expect the rotation rate of a condensation to increase as its density goes up. Lines of force emanating from the star will tend to give the cloud the same rotation as the star (the FERRARO-ALFVÉN effect). If the angular velocity of the star at any time is Ω, then within the "Kepler sphere", of radius r_k given by

$$\Omega^2 r_k^3 = GM, \qquad \ldots . (28)$$

centrifugal force is small compared with gravity, and matter will continue to flow into the star; a toroidal magnetic force is generated which effectively cancels out the increase in angular velocity of the infalling matter. But the flow into the star will not be continuous, as in the problem without rotation, for hydromagnetic waves travelling out beyond r_k will steadily build up a roughly Keplerian rotation field, in which centrifugal force and gravity approximately balance. The differential rotation twists the magnetic field, and the toroidal force generated continues to transfer angular momentum to the cloud; once an element has acquired the Keplerian rotation corresponding to its distance from the rotation axis, further gain of angular momentum causes it to move *away* from the axis, dragging the field with it. The component of gravity parallel to the axis of rotation is not opposed by centrifugal force, and so flow towards the equator continues until the further distorted field exerts a strong enough force parallel to the axis to balance gravity. The twist in the field lines is transmitted to the stellar field, which reduces the angular momentum of the star, so tending to off-set the increase in angular velocity due to contraction.

Clearly, a complete discussion of the flow of matter around the star, its effect on the field, and the outward flow of angular momentum, is a formidable task. However, rough estimates suggest that a strong cloud magnetic field is able to keep the angular velocity of a proto-star low enough for isotropic contraction to proceed, provided (1) the stellar field lines remain part of the cloud field, and do not break off and form a local field; and (2) the decrease of angular velocity at the surface of the star is transmitted rapidly enough through the whole star. The first condition cannot be verified without solving the complicated hydromagnetic problem of the matter around the star. We shall bypass the question, and concentrate on the other condition, which requires that the time of travel of a hydromagnetic wave through the

star be comparable with the time of contraction: otherwise the rotation field well within the star will be unaffected by the loss of angular momentum at the surface.

Suppose first the star is freely falling under gravity. If ρ is the instantaneous density, condition (2) above is

$$R \Big/ (H/\sqrt{8\pi\rho}) \sim 1/\sqrt{G\rho} \qquad \ldots . (29)$$

or

$$\frac{4\pi}{3} R^3 \left(\frac{H^2}{8\pi}\right) \sim \frac{GM^2}{R} \qquad \ldots . (30)$$

—i.e. the magnetic energy must be comparable with the gravitational energy. But this is just the limiting condition for a magnetic star to be gravitationally bound; moreover, the ratio of the energy of a frozen-in field to the gravitational energy is invariant under contraction (MESTEL and SPITZER, 1956). Hence we may expect a star with the maximum allowed magnetic energy to remain uniformly rotating during free-fall, and so probably to lose enough angular momentum to the cloud for contraction to proceed unimpeded by centrifugal force. If the star has reached the KELVIN-HELMHOLTZ stage of slower, radiatively-controlled contraction, condition (29) is relaxed somewhat; but even so, a small enough stellar field will be unable to prevent the build-up within the bulk of the star a large centrifugal field, even if the external field is able to transport angular momentum efficiently. Contraction perpendicular to the rotation axis is halted, and a disk is formed, with pressure balancing gravity parallel to the axis, and centrifugal force balancing gravity in the plane.

The precise value of the magnetic energy of the proto-star depends on how far the break-up of the cloud into smaller, gravitationally bound masses has proceeded, following the loss of excess magnetic energy from the cloud, e.g. by the process discussed in Section 1. We now consider those cases in which magnetic braking is unable to prevent disk formation, and inquire how the magnetic field in the disk affects its subsequent evolution. Centrifugal force is now a not mere perturbation, as in a roughly spherical mass, but is equal and opposite to the component of gravity in the plane of the disk: changes in the rotation field must therefore cause simultaneous changes in the distribution of matter, so that the magnetic field may play an essential role through its interaction with the rotation field.

We adopt the following simple model. Most of the mass is supposed concentrated in a flat cylinder of mass M_0, radius $\bar{\omega}_0$, thickness $2z_0$, and uniform density ρ_0. The condition of equilibrium in the direction of the axis yields

$$\frac{2GM_0 z_0}{\bar{\omega}_0{}^2} = 4\pi G\rho_0 z_0{}^2 = c_s{}^2, \qquad \ldots . (31)$$

where c_s is the constant velocity of sound (isothermal conditions being assumed). The equation of balance in the plane of the disk is

$$\frac{\partial V}{\partial \bar{\omega}} = -\Omega^2 \bar{\omega}, \qquad \ldots . (32)$$

where $\bar{\omega}$ is the distance from the axis, and V the gravitational potential, which can

be calculated directly from the assumed distribution of matter; the cylinder is found to rotate approximately uniformly with angular velocity Ω_0 given by

$$\Omega_0^2 = 2\pi G\rho_0 z_0/\bar{\omega}_0 = GM_0/\bar{\omega}_0^3. \quad \ldots . (33)$$

The assumption of uniform density is reasonable, as strong non-uniformities would by (32) demand a strongly non-uniform rotation field, which the magnetic field would eliminate, simultaneously smoothing out the distribution of matter. Surrounding this cylinder is an envelope of low mass, with a Keplerian angular velocity field Ω:

$$\Omega^2\bar{\omega}^3 = \Omega_0^2\bar{\omega}_0^3 = GM_0. \quad \ldots . (34)$$

The magnetic field adopted is the simplest possible one that is symmetrical about the rotation axis, and independent of height:

$$(H_{\bar{\omega}}, H_z, H_\phi) = \left(\frac{A}{\bar{\omega}}, 0, H_\phi,\right) \quad \ldots . (35)$$

where $(\bar{\omega}, z, \phi)$ are cylindrical polar co-ordinates referred to the axis and the mid-plane of the disk; A is a constant, and H_ϕ a function of $\bar{\omega}$ only. Such a field satisfies $\nabla.\boldsymbol{H} = 0$, except on the axis, where there is a uniform line distribution of magnetic poles. A more accurate theory would remove this blemish by allowing $\boldsymbol{H}$ to have a z-component, and to vary with z; however, the effects looked for are most easily illustrated by the simple field (35).

We first discuss the interaction of the magnetic and matter fields in the Keplerian envelope. If $v_{\bar{\omega}}$ is the radial velocity of the matter, the hydromagnetic equation yields

$$\frac{\partial H_\phi}{\partial t} = A\frac{\partial\Omega}{\partial\bar{\omega}} - \frac{\partial}{\partial\bar{\omega}}(v_{\bar{\omega}}H_\phi). \quad \ldots . (36)$$

The last term involves $v_{\bar{\omega}}$, which arises only through the interaction of the non-uniform rotation field with the magnetic field. As a first approximation, we neglect this term, and write

$$H_\phi = A\frac{\partial\Omega}{\partial\bar{\omega}}t = -\frac{3}{2}\frac{A(GM_0)^{\frac{1}{2}}}{\bar{\omega}^{\frac{5}{2}}}t. \quad \ldots . (37)$$

The ϕ-component of the equation of motion is

$$\rho v_{\bar{\omega}}\bar{\omega}\frac{\partial}{\partial\bar{\omega}}(\Omega\bar{\omega}^2) = \bar{\omega}\left(\frac{(\nabla\wedge\boldsymbol{H})\wedge\boldsymbol{H}}{4\pi}\right)_\phi = \frac{A}{4\pi}\frac{\partial}{\partial\bar{\omega}}(\bar{\omega}H_\phi), \quad \ldots . (38)$$

where ρ is the envelope density. The equation of continuity is

$$\frac{\partial\rho}{\partial t} = -\frac{1}{\bar{\omega}}\frac{\partial}{\partial\bar{\omega}}(\bar{\omega}\rho v_{\bar{\omega}}). \quad \ldots . (39)$$

Integrating (38) and (39) between $\bar{\omega}_1$ and $\bar{\omega}_2$, we find

$$\frac{A}{2}\left[\bar{\omega}H_\phi\right]_{\bar{\omega}_1}^{\bar{\omega}_2} = \left[2\pi\rho\bar{\omega}v_{\bar{\omega}}\Omega\bar{\omega}^2\right]_{\bar{\omega}_1}^{\bar{\omega}_2} + \frac{\partial}{\partial t}\int_{\bar{\omega}_1}^{\bar{\omega}_2}\Omega\bar{\omega}^2(2\pi\rho\bar{\omega})d\bar{\omega}. \quad \ldots . (40)$$

The first term on the right gives the net rate at which angular momentum is transported by moving matter out of the volume of unit height defined by the radii $\bar{\omega}_1$

and $\bar{\omega}_2$; the second term gives the rate of increase of angular momentum in the volume (through the increase in density); hence the left-hand side is just the net transport of angular momentum into the volume by the magnetic stresses. If $\rho = 0$, then the magnetic force must vanish, $\bar{\omega}H_\phi$ is constant, and $-A(\bar{\omega}H_\phi)/2$ is just the rate of outward transport of angular momentum along the field lines, per unit height. This is a particular case of the general formula given by LÜST and SCHLÜTER (1955).

From (34) and (38), we have

$$v_{\bar{\omega}} = \frac{(\frac{3}{2}A)^2}{2\pi\rho}\frac{t}{\bar{\omega}^3} \qquad \ldots . (41)$$

and from (39),

$$\rho = \rho_i + \frac{(\frac{3}{2}A)^2}{2\pi}\frac{t}{\bar{\omega}^4} \qquad \ldots . (42)$$

where ρ_i is the initial density of the Keplerian zone. Thus according to this model the Keplerian zone moves steadily outwards, and a vacuum develops between it and the uniformly rotating core.

The density at a particular point $\bar{\omega}$ does not sensibly alter until a time

$$t_1 \simeq (2\pi\rho_i)^{\frac{1}{2}}\bar{\omega}^2/(\tfrac{3}{2}A) = \bar{\omega}^2/\alpha^{\frac{1}{2}}, \qquad \ldots . (43)$$

say. The edge of the vacuum region at time t is defined by the position $\bar{\omega}_v$ of the element initially at $(\bar{\omega}_0)_{t=0}$; provided, then, it reaches a point $\bar{\omega}$ before the epoch defined by (43), $\bar{\omega}_v$ satisfies the equation given by (41)

$$\frac{d\bar{\omega}_v}{dt} = \alpha\left(\frac{t}{\bar{\omega}_v{}^3}\right) \qquad \ldots . (44)$$

so that

$$\bar{\omega}_v{}^4 = 2\alpha t^2 + (\bar{\omega}_0)^4_{t=0}. \qquad \ldots . (45)$$

Hence the time at which the edge of the vacuum has a particular value $\bar{\omega}_v$ is approximately

$$t_2 \simeq \bar{\omega}_v{}^2/(2\alpha)^{\frac{1}{2}}. \qquad \ldots . (46)$$

At this time ρ by (42) $\simeq \frac{3}{2}\rho_i$: the neglected term in (36) $-\partial(v_{\bar{\omega}}H_\phi)/\partial\bar{\omega} = 2A\partial\Omega/\partial\bar{\omega}$; and the ratio of the poloidal magnetic force to centrifugal force $\simeq 1/4$. For the present illustrative example, the approximations therefore remain tolerably good: by the time they begin to break down, at a particular distance $\bar{\omega}$, the vacuum has reached $\bar{\omega}$ and the theory no longer applies. (The neglect of the inertia terms in (34) may similarly be justified.)

Within the vacuum region, between the uniformly rotating core and the instantaneous position of the edge of the dispersing envelope, the magnetic field must exert a vanishing force, so that $\bar{\omega}H_\phi$ is independent of $\bar{\omega}$ and equals $B(t)$, say, where

$$B(t) = -\tfrac{3}{2}A\frac{(GM_0)^{\frac{1}{2}}}{\bar{\omega}_v{}^{\frac{3}{2}}}t = -(\tfrac{3}{2}A)^{\frac{1}{2}}(GM_0)^{\frac{1}{2}}(\pi\rho_i)^{\frac{1}{2}}t^{\frac{1}{2}}, \qquad \ldots . (47)$$

using (37) and (45).

The equation giving the rate of contraction of the uniformly rotating dense disk is

$$\frac{d}{dt}\int_0^{\bar{\omega}_0} 2\pi\rho_0 z_0 \Omega_0 \bar{\omega}^2 \bar{\omega} d\bar{\omega} = \int_0^{\bar{\omega}_0} \bar{\omega}\left(\frac{A}{4\pi\bar{\omega}^2}\right)\frac{\partial}{\partial\bar{\omega}}(\bar{\omega}H_\phi) z_0 2\pi\bar{\omega}d\bar{\omega}, \quad \ldots . (48)$$

which reduces to

$$\tfrac{1}{5}GM_0{}^2\bar{\omega}_0{}^{-\frac{1}{2}}\dot{\bar{\omega}}_0 = -\tfrac{1}{2}(Az_0)(\tfrac{3}{2}A)^{\frac{1}{4}}(GM_0)^{\frac{1}{4}}(\pi\rho_i)^{\frac{1}{4}}t^{\frac{1}{4}}, \quad \ldots . (49)$$

with the help of (31), (33), and (47).

Although for simplicity we have assumed A independent of time in discussing the envelope, in fact the freezing of the field into the disk requires

$$Az_0 = \text{constant} = (Az_0)_i \quad \ldots . (50)$$

the suffix i again referring to initial values. As A appears as a fourth root in $[\bar{\omega}H_\phi]_{\bar{\omega}_0}$, we shall treat it there as constant, while using (50) in (49). Integrating (49), we find as a characteristic time for the total loss of angular momentum by the dense disk:

$$t = (\bar{\omega}_0)_i^{\frac{2}{3}}\left[\frac{5G^{\frac{1}{2}}M_0{}^{\frac{3}{2}}}{(\frac{3}{2}A)^{\frac{1}{4}}(\pi\rho_i)^{\frac{1}{4}}(Az_0)_i}\right]^{\frac{4}{3}} \quad \ldots . (51)$$

At a point $\bar{\omega}$ within the disk, the toroidal component of the field H_ϕ is found to be

$$H_\phi = \left(\frac{\bar{\omega}}{\bar{\omega}_0}\right)^3 (H_\phi)_0. \quad \ldots . (52)$$

As developed so far, the theory gives no reason why the contracting mass should go over from the disk-like shape to the spheroid characteristic of a star with a moderate rotation rate. However, we have assumed that the disk remains isothermal —the compressional energy generated during the slow contraction is rapidly radiated away. As the density increases, the disk becomes more and more opaque, and it is probable that ultimately, even in the comparatively slow contraction determined by magnetic removal of angular momentum, the rate of generation of heat is too rapid, and a high thermal pressure is built up. At this stage, thermal pressure takes over from centrifugal force as the principal force balancing gravity. Magnetic removal of angular momentum continues, but the rate of contraction is now given by the rate of loss of energy down the equilibrium temperature gradient: the disk has become a proto-star.

Clearly any mechanism that tries to produce uniformity of angular velocity in the disk will lead to similar results; and in fact hydrodynamic turbulence has been invoked by VON WEIZSÄCKER (1944) as the means by which the envelope of a disk may be dispersed, carrying off with it excess angular momentum. Following criticisms by TER HAAR (1950), who pointed out that a wholly turbulent envelope would decay too rapidly to be effective, LÜST and SCHLÜTER (1955) have discussed the case of a magnetic star with a co-rotating KEPLER sphere, surrounded by a turbulent enveloping disk with a Keplerian rotation field. Again the turbulent envelope disperses, carrying off with it angular momentum, which is transferred from the star to the envelope by the field lines in the KEPLER sphere. In our discussion there is no KEPLER

sphere, as all the disk is in centrifugal rather than pressure equilibrium, except at the end of the process. A more important difference is that no appeal is made to turbulence. As the rotation field in the disk is such that the angular momentum per unit mass is an increasing function of ϖ, RAYLEIGH's criterion yields stability against radial perturbations. For this reason it seems better to invoke only the magnetic field as the agent by which both angular momentum and matter are redistributed.

The present discussion is not intended as more than an illustration of the sort of effect a magnetic field is likely to have. The important point is that if, within the time-scale of free-fall or KELVIN-HELMHOLTZ contraction, magnetic braking of a proto-star is too slow, then the system adjusts itself so that the time of contraction is *determined* by the time of magnetic braking. However, there are a number of serious approximations made above, which a more adequate theory should remove. First, the magnetic field considered is over-simplified; a genuine field, lying in the disk, but without fictitious sources, would change the sign of both poloidal and toroidal components between the upper and lower faces of the disk. More important, the picture of a vacuum region between a contracting central mass and a dispersing envelope is clearly untenable, as there must be continuity of pressure between the mass and envelope. What must happen is that the masses of the central mass and the envelope are not constant. As the envelope disperses, the density *at the edge* of the central mass decreases; the originally uniform angular velocity field there changes to the Keplerian field corresponding to a slightly lower central mass, so that some of the matter of the original central mass disperses with the original envelope. This process will continue until the remaining central mass has become a star. This correction is important, not only because of the lower mass of the ultimate star, but because a higher rate of braking will result; as it is, rough numerical estimates suggest that formula (51) gives too long a time-scale. Further mass loss will probably result through rotational instability during the final stage when the central mass has both a high thermal pressure and a strong centrifugal field. Finally, we have ignored the possibility of sub-condensations being formed in the disk by gravitational instability: the presence of a magnetic field is likely to assist this process, again through removing angular momentum from the condensations.

REFERENCES

BONDI, H.	1952	*M.N.*, **112**, 195.
LÜST, R. and SCHLÜTER, A.	1955	*Z. Astrophys.*, **38**, 190.
MESTEL, L. and SPITZER, L. Jr.	1956	*M.N.*, **116**, 503.
TER HAAR, D.	1950	*Ap. J.*, **111**, 179.
VON WEIZSÄCKER, C. F.	1944	*Z. Astrophys.*, **22**, 319.

Observational Aspects of Cosmology

D. W. Sciama
King's College, University of London

Summary
The article surveys the present observational status of cosmology. Despite several recent attempts, it is still not possible to decide whether the Universe is in a steady state or whether it has evolved from a much denser configuration. It is pointed out that radio astronomy will play an important role in cosmology, since radio sources can be detected at distances beyond the reach of the 200-inch telescope.

Introduction

My aim in this article is to survey the present observational status of cosmology. The reason for writing such a survey at the present time is that in the last few years several attempts have been made to obtain important new information about the large-scale structure of the Universe. It must be admitted that none of these attempts has yet been successful, but so many phenomena are now relevant to the cosmological problem that a systematic survey of the whole field may be of value.

As we want to emphasize observational questions, no detailed theoretical discussions will be given. However, one feature of the theory should be mentioned here since it provides the framework of the article. This is the distinction between the kinematical properties of the universe, which are based on very general assumptions about matter and radiation, and the physical properties, which involve Einstein's field equations and much of the apparatus of modern physics. The kinematical part of the theory is so general that it enjoys the privilege of being the least controversial part of cosmology. At the same time, it provides a useful classification of model universes in terms of which many of the observations can be interpreted. The first part of this article is therefore concerned solely with the kinematical aspects of the problem. The dynamical and physical aspects are discussed in *PART II*.

The reader who wishes to study theoretical cosmology is referred to the following detailed accounts, which also contain fairly complete references to earlier literature: Heckmann (1942), Bondi (1952, of which a new edition is promised soon), McCrea (1953), McVittie (1956), Schrödinger (1956), and Solvey Conference (1958). For less technical recent accounts of cosmology see *Scientific American* (September 1956), Munitz (1957) (a source book ranging from "Babylonian Myth to Modern Science"), McCrea (1958), and Wintrow (1959). A general discussion of the problems arising from the *uniqueness* of the universe will be given in a forthcoming book by the author (*The Unity of the Universe*, Faber and Faber, London, 1959).

PART I

Kinematical Properties of the Universe

A priori the large-scale structure of the universe could be extremely complicated. However, as in other branches of physics, the greatest progress is made by starting from the simplest assumptions that are not inconsistent with any well-established

* *Acknowledgements*—I should like to express my gratitude to the following people, who showed me their work before publication, and who gave me permission to quote from it: L. H. Aller, E. M. and

fact. We therefore begin by assuming that on a certain scale of size the universe has a simple structure. The scale of size that must be adopted for this to be even approximately true is initially unknown and must be derived from observation, but we might hope eventually to discover theoretically what determines it. We shall discuss this problem in Part II(iv); in order to make a start we shall suppose here that the basic unit of the simple structure is a cluster of galaxies, which we shall call a particle.

The simplicity assumptions that are usually made can be reduced to the following (WALKER, 1944):

Each particle always sees an isotropic distribution of particles around it.

We shall call this the cosmological principle or CP. One can deduce from the CP that all the particles will describe the system in the same way. In such a universe one can distinguish neither position nor direction, that is, one cannot tell where one is or in which direction one is looking. However, one can tell the time, for the density of particles will in general be changing, so that this density will serve as a clock.

Of course we do not expect the CP to be exactly satisfied even by our complex particles. Some statistical averaging process is needed, but once again the scale of this averaging is not determined *a priori*. Both observationally and theoretically we are here faced with a difficult question, which we return to in Part II(iv), but again in order to make a start let us assume that the CP is exactly satisfied. Then it can be shown (BONDI, 1952; MCCREA, 1953; ROBERTSON, 1956) that the metric of the universe can be written in the Robertson-Walker form:

$$ds^2 = dt^2 - \frac{R^2(t)(dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2)}{(1 + \frac{1}{4}kr^2)^2} \qquad \ldots . (1)$$

where $R(t)$ (the expansion factor) is an arbitrary function of time and k (the spatial curvature in units of R^{-2}) has the possible values[1]

$k = +1$ elliptic spatial universe
$k = 0$ Euclidean spatial universe
$k = -1$ hyperbolic spatial universe

With this form of the metric the co-ordinates of a particle (r, θ, q), are independent of t (co-moving co-ordinates), and the paths of light-rays are given as usual by $ds^2 = 0$. Furthermore, for a given particle $dr = d\theta = dq = 0$, so that $dt = ds$, which means that t measures the proper-time for each particle.

Before discussing the observational significance of the Robertson-Walker metrics, it is convenient to mention here the effect of making one further simplifying assumption. This consists of adding to the CP the following statement:

Each particle always sees the ***same*** *distribution of particles around it.*

This is known as the perfect cosmological principle or PCP (BONDI and GOLD,

G. R. BURBIDGE, A. G. W. CAMERON, D. O. EDGE, G. B. FIELD, W. A. FOWLER, T. GOLD, D. S. HEESCHEN, F. HOYLE, D. LAYZER, A. C. B. LOVELL, B. Y. MILLS, G. MÜNCH, J. NEYMAN, M. RYLE, P. A. G. SCHEUER, E. L. SCOTT, J. R. SHAKESHAFT, A. E. WHITFORD.

I should like to thank RIAS for their support during the summer of 1957, when I began preparing this article. I am also grateful to B. E. J. PAGEL and F. A. E. PIRANI for their valuable criticisms of the first draft of the manuscript.

1. (Added in proof.) This classification is not correct, as there are other possibilities for the topology of space. For instance, the spatial universe might be closed even if $K = -1$.

1948). In this more restricted type of universe one can no longer tell the time. The number of possible metrics is now drastically reduced. The only possibilities are.

(i) $R = \text{const.}\, k$ is arbitrary. (Einstein Universe) and
(ii) $R = e^{t/\tau_0}, k = 0$. τ_0 is a constant. (de Sitter Universe). . . . . (2)

We now consider the attempts that have been made to test the Robertson-Walker formula (1). No very precise observational study has yet been made of the assumptions underlying (1), since statistically adequate data do not exist. (However, several extensive galaxy counts are currently being made [ABELL, 1957, 1958; SHANE, 1955; ZWICKY, 1957] which may be important for this problem). Because of this, the usual approach has been to accept the Robertson-Walker models, and to try and determine the parameters k and $R(t)$. Of course, one must first derive from (1) theoretical relationships between directly observable quantities. This has been done by ROBERTSON (1955, 1956), whose work we follow.

The first theoretical relationship we shall discuss is that between the apparent bolometric magnitude of an *average* galaxy and its red-shift z ($= \delta\lambda/\lambda$), deduced from the properties of light-propagation in the metrics (1). The result is

$$m_{bol} = M_0 - 45{\cdot}06 + 5 \log \tau z + 1{\cdot}086(1 + q - 2\mu)z + O(z^2) \quad . . . . (3)$$

where $\tau = R/\dot{R}$ evaluated at the time of observation
$q = R\ddot{R}/\dot{R}^2$ evaluated at the time of observation
μ relates to possible intergalactic absorption and also to a possible evolution of the absolute magnitude of an average galaxy.
M_0 is the absolute magnitude of an average galaxy.

The first term in z gives the well-known linear Hubble law. For small z, the concept "distance" is unambiguous, and we can write

$$z \sim \frac{v}{c} \sim \frac{\text{distance}}{c\tau}$$

where v is the velocity of recession. The observed value of τ is about $6{\cdot}10^9$ years, but this value may well have to be revised upwards (BAADE, 1956; EGGEN, GASCOIGNE and BURR, 1957; SANDAGE, 1958). Hence the present value of $\dot{R}/R$ is known approximately. Of course the non-zero value of $\dot{R}$ rules out the static Einstein universe $R = \text{const.}$, so that if the PCP is correct, the universe must conform to the de Sitter metric (2). This is the metric of the steady-state universe (BONDI and GOLD, 1948; HOYLE, 1948, 1949; BONDI, 1952; MCCREA, 1950, 1951, 1953) in which the density is maintained at a constant value by the continual creation of matter.

If (3) is to be used to discriminate between the two versions of the cosmological principle, one must determine the value of the acceleration q from the second term in z. Its theoretical value for the PCP is 1 (since $R = e^{t/\tau}$ in that case) whereas for the CP it can have any value. Now the linear Hubble law appears to be a good approximation to the observations for values of z lying between 0 and about 0·1. Thus, observations must be made in the region $z > 0{\cdot}1$, where the apparent magnitudes of galaxies are very faint. However, by pushing photographic techniques to their limit, HUMASON, MAYALL and SANDAGE (1956) attempted to test (3) out to $z \sim 0{\cdot}2$, and so obtain a value for q. More recently BAUM (1957) has used photoelectric methods to extend the observations out to $z \sim 0{\cdot}4$.

Unfortunately, for these large values of z selection effects become appreciable (Scott, 1957; Stock and Schuecking, 1957; Sandé, 1957). The main effect arises from the unavoidable selection of galaxies at great distances whose absolute brightness tends to be greater than average. This leads to an *underestimate* of q. Miss Scott attempted to compute this error, but the computation depends critically on the luminosity function for galaxies, and also on the distribution of cluster sizes. These are not well-known at the moment, so that a reliable value for q cannot be obtained.

In order to illustrate the extent of the uncertainty in q, let us first consider the results obtained when selection effects are neglected. After introducing various corrections e.g. for converting photographic magnitudes into bolometric magnitudes (the K term), and putting $\mu = 0$ (in the absence of any evidence that μ differs appreciably from zero), Humason, Mayall, and Sandage obtained the results

$$\tau = 5{\cdot}4 \,.\, 10^9 \text{ years}$$

$$q = -2{\cdot}5 \pm 1$$

The value of τ has the same uncertainty as the distance scale (cf. Baade, 1956; Eggen, Gascoigne and Burr, 1957; Sandage, 1958) since it involves the absolute magnitude M_0, and a new attempt is being made to improve it.

It is clear that the uncorrected value of q differs substantially from the value 1 implied by the PCP. Baum's uncorrected results show a similar discrepancy with the PCP. He obtains a larger value of q, but his estimated error is smaller in consequence of the greater accuracy of photoelectric techniques and the larger values of z involved. His uncorrected value of q is:

$$q = -1 \pm 0{\cdot}5$$

Now, as we have already pointed out, the Scott correction will further increase q. The amount of this increase is uncertain, but by choosing plausible values for the luminosity function and the distribution of cluster sizes, one can correct Baum's result to $q > +1$ (cf. Scott, 1957, Fig. 8). It is clear then that the Scott correction will have to be more reliably estimated before the value of q can be used to distinguish between the two cosmological principles.

A promising alternative is offered by Morgan and Mayall's (1957) suggestion that there may be a correlation between the line-widths in the integrated spectrum of a galaxy and its absolute magnitude. Such a correlation would enable one to determine the distance of a galaxy without any selection effects.

Unfortunately, even if the red-shift data were accurate enough to determine q, they would not determine the sign of the curvature k, since k only occurs in the term $O(z^2)$, which cannot be reliably estimated from observations out to $z \sim 0{\cdot}4$. However, there exists another theoretical relationship between observables which has been used to try and determine k. This involves number counts of galaxies of different apparent magnitudes m. If $N(m)$ is the number of galaxies out to a given apparent magnitude m, then the theoretical relationship is (Robertson, 1955)

$$\frac{1}{N}\frac{dN}{dm} = 1{\cdot}382\left\{1-(1-\mu)z+\left[\frac{3}{2}+\frac{k}{5\dot{R}^2}+\frac{q}{2}(1+\mu)-\frac{7}{2}\mu+\mu^2-K'\right]z^2\right\}+O(z^3)$$

where K' involves the second time derivative of M_0. This relation has to be modified for the steady state model, to allow for the continual formation of new galaxies.

In his now classical work, HUBBLE (1936) attempted to determine k from this formula. However, the presence of the unknown q in the coefficient of z^2 makes this very difficult. Furthermore, the relation is sensitive to random fluctuations in the distribution of galaxies. ROBERTSON suggests that the only way it could be used is for determining μ empirically from the first-order term in z.

Recently a new attempt has been made to obtain cosmological information from number counts—but by using radio sources rather than optical galaxies (RYLE and SCHEUER, 1955. SHAKESHAFT *et al.* 1955; MILLS and SLEE, 1957; EDGE, SCHEUER and SHAKESHAFT, 1958). Of course, one has no red-shift data for these sources and no distance determinations for the majority of them, so that it must first be established indirectly that these objects are extragalactic! Some of the optically identified sources are known to be extragalactic (MINKOWSKI, 1958), but these may form a small percentage of the 2,000 or so observed in the first large-scale survey (SHAKESHAFT *et al.*, 1955). However, these sources are distributed isotropically, so that if they are not extragalactic they must be close enough to the sun for the anisotropy of the Milky Way to be concealed. For the purposes of discussion it will be supposed that most of the sources are extragalactic (*cf.* RYLE 1958), but it should be borne in mind that this has not yet been definitely established.

Since the radio astronomers work with the number of sources per unit solid angle $N(I)$ whose apparent intensity exceeds I, we must express the theoretical relationship in terms of these variables. For evolving models of the universe, that is, those which satisfy CP but not PCP, the result is

$$(\text{CP}) \quad \log N = C_0 - \frac{3}{2}\log I - \left(\frac{3}{\tau}+\mu\right)KI^{-\frac{1}{2}} + \left(\frac{12}{\tau^2}+\frac{3}{10}\frac{k}{R^2}+\frac{3}{2}\frac{q}{\tau^2}+\mu'\right)K^2I^{-1} + O\left(I^{-\frac{3}{2}}\right)$$

where C_0 is a constant involving the number density of sources and their absolute intensity I_0. K is a constant involving I_0. μ involves the first time derivative of I_0, and μ' the first and second time derivatives of I_0.

The first-order deviation from the static Euclidean result is given by the term

$$-\frac{3}{\tau}KI^{-\frac{1}{2}}.$$

This *reduces* the value of N from its Euclidean value. It arises from the Doppler shift which decreases the apparent intensity of each source. Hence for a given apparent intensity, one is seeing less far into space than in the static case, and so seeing a smaller number of objects.

An effect in the opposite direction has also been suggested (e.g. RYLE, 1958), namely, that if a substantial number of sources are colliding galaxies (like Cygnus A) then the number of sources *increases* with distance, since collisions were more frequent in the past when the galaxies were closer together. However, most collisions occur within a cluster of galaxies (BAADE and SPITZER, 1951). These do not share in the expansion, so that this effect would only operate at distances where clusters are nearly in contact with one another.

The steady-state model contains fewer sources per unit volume in the past than do the evolving models. This leads to a still greater reduction from the static case. The result is

$$(\text{PCP}) \qquad \log N = C_0 - \frac{3}{2}\log I - \frac{21}{4\tau}KI^{-\frac{1}{2}} + \frac{35}{\tau^2}K^2I^{-1} + O(I^{-\frac{3}{2}})$$

The coefficient 35 in the second-order term is the result of rounding off a complicated fraction by about 1 per cent. In this case, of course, there will be no terms in μ and μ'.

Unfortunately, these equations cannot be used directly to interpret the observations. The reason is that they are derived on the assumption that all the sources have the same absolute intensity I_0. However, radio sources are known to vary in absolute intensity over a far greater range than optical galaxies. Hence a given apparent intensity corresponds both to nearer fainter sources and further brighter ones (which will be more affected by relativistic corrections). Since the luminosity function is not at all known, it will be very difficult to distinguish its influence from that of τ, q and k on the one hand, and $\dot{I}_0$ and $\ddot{I}_0$ on the other.

We now consider the observations that have so far been made. As we shall see, they are in conflict with one another, so that it is not worth attempting to interpret them in terms of the theoretical relationships. In the first large-scale survey (Ryle and Scheuer, 1955; Shakeshaft *et al.*, 1955) nearly 2,000 sources were measured at a frequency of 81·5 Mc/s, by means of a Ryle-Hewish (1955) double interferometer. The plot of $\log N$ against $\log I$ showed a substantial *positive* deviation from the static Euclidean relation $\log N = C - \frac{3}{2}\log I$, that is, for small I, N exceeded the static Euclidean value—indeed for small values of $\log I$, the slope was ~ -3.

However, it was recognized that number counts of sources would not be reliable since the instrument used is confusion-limited, that is, its recorded output consists of the overlapping traces of a large number of sources. Under these conditions it is difficult to infer actual numbers of sources from the observations. For this reason, Scheuer (1957) developed a statistical method of analyzing directly the deflexions of the receiver. For given assumed distributions of radio sources, he calculated the number $P(D)$ of deflexions of size D. He concluded from this analysis that there was a deficit of bright sources or an excess of faint ones as compared with the static Euclidean case. However this result might be spurious, since the random errors in the measured intensities probably increase with decreasing intensity (Bolton, 1956a, b), which will lead to an apparent excess of faint sources (cf. Bolton, Stanley and Slee, 1954).

The second large-scale survey has been made by the Australian group (Mills and Slee, 1957; and Mills *et al.* unpublished), who used a different type of interferometer—a Mills cross (Mills and Little, 1953). This has a pencil beam reception pattern which reduces the effects of confusion. The complete Australian results will shortly be published, but a preliminary account has already been given (Mills and Slee, 1957). There is marked disagreement with the Cambridge results, since Mills and Slee obtain the relation

$$\log N = C_0 - 1{\cdot}8 \log I$$

The difference between 1·8 and the static Euclidean value 1·5 is not regarded as significant, and is attributed to instrumental effects. The complete survey confirms this preliminary result (Mills, private communication).

Further discrepancies arise when the sources given by the two surveys are compared in the region of overlap. Indeed the two surveys are almost completely discordant. The situation is made still worse by the second Cambridge survey that has recently been completed (at 159 Mc/s), (Edge, Scheuer and Shakeshaft, 1958), for the sources of this survey are discordant with those of both the previous ones! In addition, the direct analysis of deflections is more unreliable for this survey,

since the Cambridge receiver has a bias against large-diameter sources, which is more important at the higher frequency. (Hazard and Walsh, 1958; Pawsey 1958; Ryle 1958).

These inconsistencies will no doubt be removed by further investigations. Of particular interest is the new Cambridge radio telescope, which has greater resolution and sensitivity than any previous instrument, and which should give reliable results for the faint sources. If such results could be obtained, they might have cosmological value despite the complexities of the theoretical relations. For instance, certain theoretical possibilities might be ruled out if they could fit the observations only with unreasonable assumptions about I_0 or the luminosity function.

It is possible that radio astronomy will make a more fruitful contribution to cosmology through observations of the 21 cm line of neutral hydrogen, since one then has a measurement of the all-important red-shift. Recently several galaxies and clusters of galaxies have been detected at 21cm. Those of direct interest to cosmology are Cygnus A (two colliding galaxies seen in absorption) (Lilley and McClain, 1956, 1958), and the clusters in Coma, Hercules and Corona Borealis (seen in emission); (Heeschen, 1956, 1957; Heeschen and Dieter, 1958.). The radio red-shifts are all in good agreement with the optical values. This shows, incidentally, that $\delta\lambda \propto \lambda$ over a range of λ of 5.10^5 to 1 (cf. Minkowski and Wilson, 1956) in agreement with relativistic theory, but in disagreement with some other explanations of the red-shift (cf. Whitrow, 1954).

The potentialities of the 21 cm method are immense. For instance, Heeschen estimates that with a 140-foot telescope a cluster like Hercules would be observable at distances exceeding 3.10^9 light years ($z \geqslant 0{\cdot}5$), a region for which different cosmological models give substantially different predictions. Furthermore, one can improve the sensitivity of the equipment by using the recently developed solid-state microwave amplifiers (masers).

It may also be possible to detect the integrated (red-shifted) 21 cm radiation from all the distant galaxies. If this could be done, it would discriminate very directly between the two cosmological principles, for according to the CP the spectrum is independent of frequency near the cut-off at 1420 Mc/s (21 cm), whereas according to the PCP it depends on the cube of the frequency. This striking difference should be easy to detect, if only the radiation can be observed at all in the presence of the background continuum radiation (e.g. by switching through the cut-off at 1420 Mc/s).

PART II

Physical Properties of Model Universes

Since the kinematical considerations of the last section do not yield much detail about the structure of the universe, we now turn to Einstein's field equations and the results of terrestrial physics to try and obtain more information. The use of Einstein's equations would enable us to infer much more from a reliable red-shift-apparent magnitude curve, while the introduction of the general range of physics enables us to calculate (rather roughly, of course) some of the physical properties of the different models, and to compare these properties with observation.

(i) *The Red-shift Apparent Magnitude Data*

ROBERTSON (1955, 1956) and HOYLE and SANDAGE (1956) have given useful summaries of the relevant implications of EINSTEIN's field equations. We shall consider these implications separately for the CP and the PCP.

(a) CP

Einstein's equations in the usual notation are

$$G_{\mu\nu} - \tfrac{1}{2}Gg_{\mu\nu} + \lambda g_{\mu\nu} = -\kappa T_{\mu\nu}$$

The discussion of their implications is much simplified when the cosmical constant λ is assumed to be zero, so let us first consider that case.

(a_1). $\lambda = 0$.

When the Robertson-Walker metric is inserted into the field equations, one obtains the following relations:

$$\frac{\dot{R}^2}{R^2} + \frac{2\ddot{R}}{R} = -\frac{k}{R^2}$$
$$\frac{\dot{R}^2}{R^2} - \frac{8\pi G\rho}{3} = -\frac{k}{R^2} \qquad \ldots\ldots (4)$$

where G is the Newtonian gravitational constant, ρ is the density of matter, and R and its derivatives are evaluated at the moment of observation. In these relations the pressure arising from random motions of the clusters of galaxies, and from electromagnetic radiation, has been neglected. This seems reasonable, except perhaps in the early stages of evolution.

These equations are easily cast into the more convenient form

$$\frac{\ddot{R}}{R} = \frac{q}{\tau^2} = -\frac{4\pi G\rho}{3} \qquad \ldots\ldots (5)$$

$$\frac{1}{\tau^2}(2q+1) = -\frac{k}{R^2} \qquad \ldots\ldots (6)$$

These relations show that, in contrast to the kinematical discussion, q and k are no longer independent.

If ρ could be measured accurately, (5) would determine q (since τ is known). Insertion of this value of q in (6) would then give the sign of k. Unfortunately, ρ is not known very accurately. The masses of galaxies (themselves not known very well) lead only to a lower limit for ρ. However, the previously mentioned detection of neutral hydrogen in clusters of galaxies, and the radio astronomical search for intergalactic matter (see (ii)) suggests that a more reliable value for ρ may eventually be forthcoming.

In the absence of this information, one must use the value of q determined from the red-shift-apparent magnitude curve to obtain k from (6). If $q < -\frac{1}{2}$, $k = +1$, and the spatial universe is closed and cyclic. If $q = -\frac{1}{2}$, $k = 0$ and the spatial universe is flat about the structure of the universe, we now turn to EINSTEIN's field equations and (Einstein—de Sitter model). If $-\frac{1}{2} < q \leqslant 0$, $k = -1$ and the universe is open (see

VOL. 3 1960

footnote on p. 312) and indefinitely expanding. If $q > 0$, (5) breaks down and we must have a non-zero cosmical constant (or possibly the steady-state model if $q = 1$). This range of possibilities shows how important the SCOTT correction is, because it might well swing q from -1 (BAUM) to > 1, that is, through the whole range of possible types of model!

(a_2). $\lambda \neq 0$

There are now many more possibilities. We shall restrict ourselves to those models that are in the singular state $R=0$ at $t=0$ ("explosive models"). This excludes the EDDINGTON model which starts out from the (unstable) EINSTEIN configuration. This model can hardly be distinguished observationally from one of the explosive models which passes slowly through the Einstein state (LEMAÎTRE). The various possibilities are illustrated in Fig. 1, (after ROBERTSON, 1956). If one could estimate

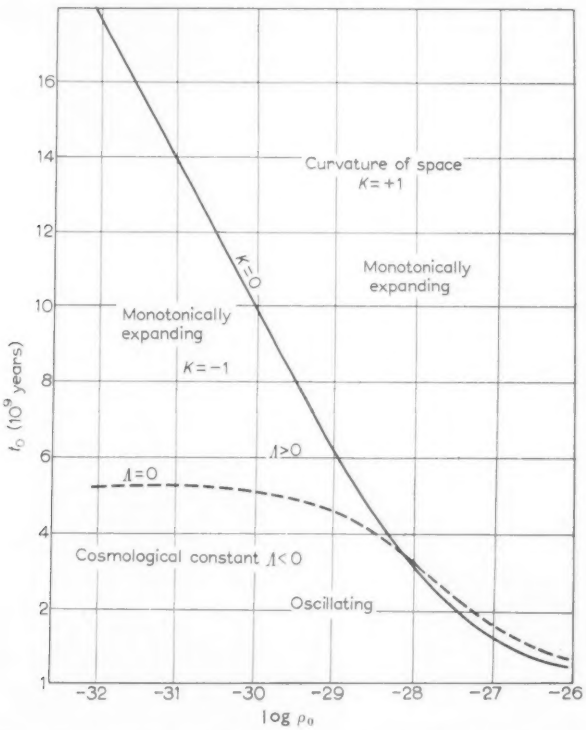


FIG. 1. Dependence of the physical model of the Universe on the two parameters density ρ_0 and epoch t_0. The represented time-density relation is based on a value of $\tau = 5{\cdot}4 \times 10^9$ years for HUBBLE's constant.

ρ and t (the present epoch since the beginning of the expansion), then the diagram shows the corresponding values of λ and k. The $\lambda = 0$ curve should probably be higher, since τ is probably greater than $5{\cdot}4.10^9$ years (BAADE, 1956; EGGEN *et al.*, 1957; SANDAGE, 1958). If one estimates $\rho \sim 10^{-29}$g/cc and $t \sim 6.10^9$ years, (from the ages of star clusters (HOYLE and HASELGROVE, 1957, SANDAGE, 1957)) with $\tau \sim 6.10^9$ years, one is in the region of the diagram where both λ and k might be positive, zero or negative.

Alternatively, if q were known from the red-shift curve, then the equation

$$\frac{q}{\tau^2}+\frac{4\pi G\rho}{3}=\frac{\lambda}{3}$$

gives a relation between λ and ρ which would restrict the range of possibilities given by the Robertson diagram. However, one would not be led to a unique solution. On the other hand, if q turns out to be unity, one would be tempted to adopt the steady-state model, although there exist values of λ which also lead to this result. Only an independent estimate of ρ or t could distinguish between these possibilities.

(b). PCP

Independently of any field equations we know that the PCP is characterized by $k=0$ and $q=1$. However, just as with the CP, we can obtain a relation involving the density if we invoke field equations. HOYLE (1948, 1949) and MCCREA (1951) have shown how to adapt EINSTEIN's equations to the PCP. In this case there can be no cosmical constant, and one obtains the relation

$$\frac{8\pi}{3}G\rho\tau^2=1$$

so that this model leads to a definite value for the density. For $\tau\sim 6.10^9$ years, $\rho\sim 5.10^{-29}$ g/cc. The creation rate can also be deduced—it is $3\rho/\tau$, that is, 10^{-45} g/cc/sec., but of course there is no hope of observing this directly. On the other hand, as we shall see in the next section, it may be possible to determine the density ρ from observation.

(ii) *Intergalactic Matter*

As we have just seen, a reliable value for the mean density of matter in the universe would be of great help in discriminating between the various possible models. Is there any hope of obtaining a sufficiently accurate value of ρ from observation?

Until recently, the answer would have been a definite "no". The observed masses of galaxies, themselves very uncertain, give only a lower limit to ρ. This lower limit is generally considered to be around 10^{-30} g/cc, which is somewhat too low to be of direct cosmological relevance. Of course, an upward revision of these galactic masses and the discovery of many faint dwarf galaxies (ZWICKY, 1957) might substantially increase this lower limit. However, the possibility that should be mentioned here is that intergalactic matter may have a density exceeding that of the galaxies—and that this density may be measurable.

The first evidence for intergalactic matter came from ZWICKY's (1957) discovery that some pairs of galaxies are connected by luminous cords. However, these cords may not be typical of conditions in intergalactic space, since they may be the result of some dynamical interaction between relatively close pairs of galaxies. An attempt to detect the general distribution of intergalactic hydrogen has been made by FIELD (1958) (cf. WHITFORD, 1954a, b; MCCLAIN, 1956; LILLEY and MCCLAIN, 1958). FIELD used the 60-foot Harvard radio telescope to study the spectrum of Cygnus A, in the hope of finding intergalactic absorption in the 21 cm region. In this first attempt

there was no detectable absorption, corresponding to an upper limit of 1 per cent. This implies that

$$\frac{n_H}{T_s} < 10^{-6}$$

where n_{H} is the number density of neutral hydrogen atoms, and T_s is the spin temperature in degrees Kelvin, which specifies the relative occupation numbers of the singlet and triplet hyperfine levels of hydrogen.

In order to derive n_{H} from this result, one must obtain an estimate of T_s. FIELD has studied theoretically the various collision processes that determine T_s; (PURCELL and FIELD, 1956; FIELD, 1958 and 1959). The discussion is not quite straightforward since the effect of collisions on T_s itself involves n_{H}. In addition, ultraviolet radiation from galaxies could increase T_s considerably. FIELD's tentative conclusions are:

(a) If there is only a few per cent ionization the total density of neutral hydrogen is less than $5 \cdot 10^{-29}$ g/cc.

(b) If the ionization is significant (perhaps 10 per cent or more), the spin temperature T_s will be mainly determined by the line-shape of the Lyman α radiation arising from recombinations. This shape is probably determined by the kinetic temperature, whose value we do not know. In other words, if the ionization exceeds 10 per cent, no definite conclusions can be drawn from FIELD's observations.

From the cosmological point of view, a substantial degree of ionization would not be surprising. The reason for this is that the lifetime of a proton against recombination in intergalactic conditions is of the same order as HUBBLE's constant τ over a wide range of kinetic temperatures (even in the absence of ultra-violet radiation) (cf. also MCCREA, 1958). Hence if matter is ionized, either as a result of the high temperatures that prevail in the early phases of the explosive models or as a result of continual creation, it will remain ionized for a "cosmological" length of time. It is possible that the proton lifetime is such that fluctuations in the density of neutral hydrogen may act as a marker for special localized conditions, e.g. of low temperatures or high densities. In this connection it is worth noting that recent 21 cm observations of M 33 (DIETER, 1958) suggest that there is a measurable amount of neutral hydrogen far outside the optical limits of the Galaxy. This material does not appear to share the rotation of the Galaxy, but rather to be in an independent state of motion. Perhaps it can be interpreted as intergalactic gas which has recombined as a result of the density increase induced by the gravitational action of M 33; (SCIAMA 1955).

If most of the intergalactic hydrogen is indeed ionized, it will unfortunately be very difficult to detect. Even out to a distance of a billion parsecs the emission measure would only be of the order of 10 (for $n_{\mathrm{H}} \sim 10^{-4}$). However, it might be observable from an artificial satellite (cf. BAUM, 1956). Even if this is not possible, FIELD's work and the Harvard radio observations suggest that we stand at the threshold of a new branch of astrophysics—the study of intergalactic matter.

(iii) *The Age Distribution of Galaxies*

The CP and the PCP differ substantially in their predictions about the age distribution of galaxies. The reason for this is that in theories based on the CP most

galaxies were formed at about the same time (see (iv)). Hence, according to such theories, the more distant a galaxy is the younger it will appear to be (because a greater light-time is involved). On the other hand, the PCP implies that the galaxies have an age distribution that is the same at all times and therefore at all distances. (This age distribution is actually exponential with an average of $\tau/3$). This difference would enable one to distinguish between the two cosmological principles, if only the ages of galaxies could be determined.

Unfortunately this cannot yet be done, although developments in our understanding of stellar evolution, (e.g. HOYLE and HASELGROVE, 1957; SANDAGE, 1957; MORGAN and MAYALL, 1957) suggest that it may soon be possible. In addition, there is a little evidence that some galaxies have been recently formed. HARO (1956a, b) and MÜNCH (1957, 1958) have found several very blue extragalactic objects which contain early-type stars, but are mostly uncondensed gas. It would appear that these systems are either very young, or that for some reason star formation has only recently become possible in them. However, if such galaxies are rare, theories based on the CP cannot be rejected since even in them *some* galaxies may have been formed recently.

Although galaxies cannot be directly dated, their properties might be used indirectly to discriminate between the two cosmological principles. For according to the PCP, no *intrinsic* property of a galaxy can change in a systematic way with distance from us. If such a change could be discovered it would disprove the PCP, even though we do not know the relation between such a property and the age of a galaxy. On the other hand, the absence of any such property is not very informative, since galaxies may not change much in the 2.10^9 years which we have looked into the past (radio sources possibly excepted!).

In 1948 STEBBINS and WHITFORD suggested that the intrinsic colour-index of a galaxy might be one such property. Their two-colour measurements of various galaxies appeared to imply that the more distant galaxies were *intrinsically* redder than nearby ones (that is, after correcting for the effects of the red-shift). On the explosive theories this would mean that younger galaxies are redder than older ones, perhaps as a result of containing more red giants (although this would imply that these stars evolve surprisingly rapidly). On the other hand, the steady state model would be ruled out.

The Stebbins-Whitford effect has such important consequences both for cosmology and for stellar evolution that it is essential that it be reliably established. However, both DE VAUCOULEURS (1948) and BONDI, GOLD and SCIAMA (1954) pointed out that there were too many uncertainties in the existing data for reliable inferences to be drawn. The most dangerous step was to use six-colour observations of M 32 to calculate the effect of red-shift on colour index, since M 32 might be atypical.

Recent work by WHITFORD and CODE (1957)—see also ZWICKY (1957, p. 219)—has shown that the use of M 32 did indeed lead to error, and that the Stebbins-Whitford effect does not exist. Distant galaxies have now been measured in six colours, so that the effects of red-shift can be directly estimated. These measurements will be of the greatest importance for galactic studies; in particular it will give us information about the spectra of galaxies in the ultra-violet, which will improve the estimate of the K term.

Now that the Stebbins-Whitford effect has been withdrawn, there is no known instance of an intrinsic property of a galaxy that changes systematically with

distance[1]. However, as we said before, this does not enable us to discriminate between the CP and the PCP, unless this result can be extended to very distant galaxies.

(iv) *The Formation and Distribution of Galaxies*

The derivation of the Robertson-Walker models is based on the assumption that the Universe is spatially homogeneous. This homogeneity is taken to be exact for mathematical convenience—in practice it has only a statistical significance. The statistical version of the assumption has never been precisely stated: it is usual to refer to averages over a "sufficiently large region" (cf. NEYMAN and SCOTT, 1958a, b). However, this region may be so large that one will have to distinguish between the spatial universe at one time (the world-map) and the universe as we see it on our past light-cone (the world-picture). The homogeneity refers to averages taken over regions in the world-map, whereas observational tests of homogeneity involve the world-picture. Hence, the assumption of homogeneity may be difficult to test (except in the steady state model, where averages in the world-picture and the world-map are the same). Furthermore, there is nothing in the theory at its present stage of development to indicate how large a region must be sampled in order to make an adequate test.

Thus on purely theoretical grounds it is important to try and calculate the properties of the "local" irregularities that can form in the universe. Of course, there is no need to stress its importance from the observational point of view: these irregularities (presumably clusters and may be super-clusters of galaxies) are, as it were, the common meeting-ground between astrophysics and cosmology.

There have been various attempts to discuss the formation of galaxies in evolving models of the universe (LEMAÎTRE, 1949; GAMOW, 1949, 1953; LAYZER, 1954). According to these theories, there was a special period of time when physical conditions in the universe were favourable for the occurrence of this process. Unfortunately, these theories suffer from the difficulty that small perturbations of the homogeneous models will not grow into galaxies in the times available (BONNOR, 1957a, b), so that large perturbations have to be invoked. This in turn leads to further difficulties, one technical and one of principle. The technical difficulty is that no one knows how to discuss gravitational condensation in the presence of the highly turbulent compressible motion that is needed to provide large density fluctuations. The difficulty of principle is that the characteristics of this violent motion (e.g. the Mach number, or the spectrum of the turbulence) are not determined by the theory, but are implicit as initial conditions in the state of the universe at $t \sim 0$. These difficulties, while not sufficiently serious to dispose of the evolving models, have stood in the way of any detailed calculations of the properties of galaxies which could be compared with observation.

The situation is different in the steady state theory, where both difficulties can be avoided (SCIAMA, 1955). It is true that one still needs a large local perturbation in the density (by a factor ~ 4), but this can be achieved without violent compressible turbulence, because in a steady state galaxies must exist at all times (or never!). These pre-existing galaxies will gravitationally perturb the intergalactic gas in a *systematic* way, and so produce a large density increase without at the same time

[1] The recent work of JUST (1959) suggests that the contents of clusters of galaxies changes systematically with distance. However, JUST has not included in his analysis all the known possible sources of error; he hopes to rectify this omission (private communication).

producing violent turbulence. The usual Jeans-criterion (JEANS 1928) can then be used in an approximate way to decide whether gravitational instability will occur in these dense regions.

The difficulty about initial conditions is eliminated by the requirement that the system be in a steady state. This requirement is so exacting that it leads to a determination of the average properties of the system of galaxies without the introduction of any arbitrary parameters. It should be emphasized that the calculations so far made do not necessarily take into account all the relevant factors. For instance, magnetic fields have been ignored, whereas since the intergalactic gas is probably substantially ionized, as shown in (ii), they may have an important dynamical effect on the condensation process; see MESTEL and SPITZER (1956) for the corresponding stellar problem. The aim of the calculations is thus not to give a complete discussion of galaxy formation, but rather to describe a model which shows how the steady state requirement can lead to definite results for the average properties of galaxies. These calculations should be extended in two directions. Firstly, all the important factors should be included, and secondly, the distribution of galactic properties about their average values should be studied.

Despite the crudity of the calculation, its implications are in fair agreement with observation. The main theoretical results are:

Average mass of a galaxy:	2.10^{43} g or 6.10^{44} g (bimodal distribution) (cf. HOYLE, 1953)
Average distance between galaxies:	4 to 10.10^{23} cm.
Peculiar velocity of a single galaxy:	4.10^7 cm/sec.
Maximum number of galaxies in a cluster:	10^4
Relative number of clusters containing $n(< 10^4)$ galaxies:	α/n^2, (α a known constant)
Average number of galaxies per cluster:	10
Percentage of galaxies that are single:	10

Thus, in a sphere of radius 10^8 light-years there will be 100,000 clusters containing altogether 10^6 galaxies, the number of clusters containing 1,000 or more galaxies is about 100, and the number containing 10^4 is about 10. None of these results conflict with the rather uncertain observational data.

Statistical studies of the observed distribution of galaxies have also been made quite independently of any particular cosmological model. (SHANE, 1955; NEYMAN and SCOTT, 1955, 1956, 1959a, b; ZWICKY, 1956, 1957; HERZOG, WILD and ZWICKY, 1957; ABELL, 1957, 1958; LAYZER, 1956; LIMBER, 1957). Various problems have been tackled, e.g. the properties of individual clusters, the size distribution of clusters, and super-clustering (cf. also DE VAUCOULEURS, 1956, 1958). Although this work has not yet led to any reliable conclusions, its further development is of vital importance in securing the framework for cosmological theorizing.

(v) *The Formation and Distribution of the Elements*

The first attempt to relate the formation of the elements to cosmological considerations was made by GAMOW in 1946. He was obeying EDDINGTON's famous injunction "go and find a hotter place" (EDDINGTON, 1926, p. 301.) GAMOW's hotter place consisted of the Universe as a whole at times near $t \sim 0$. He suggested that at those times the density and temperature were suitable for the formation of the elements from hydrogen (or neutrons) as a result of thermonuclear reactions. This led to the so-called α-β-γ theory of element formation. (ALPHER, BETHE and GAMOW, 1948). The state of this theory and its several variants was reviewed in 1950 (ALPHER and

Herman, 1950), in 1953 (Alpher and Herman, 1953) and in 1954 (*Les processus nucléaires dans les astres*, 1954, Part I).

The aim of the α–β–γ theory was to find a set of conditions of density and temperature consistent with Einstein's field equations, which, in conjunction with known nuclear reaction rates, would lead to the observed abundance distribution of the elements. Despite the freedom permitted by Einstein's equations, however, the programme soon ran into difficulties. The most direct difficulty arose from the instability of nuclei of atomic weights 5 and 8. This instability acts as a bottle-neck in the building-up process, and, despite several attempts to avoid it, it still appears to prevent the formation of an appreciable number of elements heavier than helium.

Furthermore, indirect evidence has been discovered that element formation has probably occurred since the Milky Way was formed (in addition to the conversion of H to He in main sequence stars). We may mention here Merrill's (1952) discovery of Tc in S-stars. Tc is the "missing element" (atomic number 43), and its longest lived isotope has a half-life of $\sim 2.10^5$ years. This Tc must have been formed in the S-stars themselves; (Greenstein, 1954). Further evidence comes from the fact that young stars appear to have a higher heavy element content than old stars (e.g. M. Schwarzschild, B. Schwarzschild, Searle and Meltzer; 1957), which indicates that the interstellar concentration of heavy elements increases with time.

These considerations suggest the adoption of a different approach, in which one attempts to account for the abundance distribution of the elements entirely in terms of nuclear processes in the various known types of star. The steady state model, of course, absolutely requires this point of view. Early studies along these lines were made by Farkas and Harteck (1931) and Pokrowski (1931). Recent developments in our knowledge of nuclear data and in our understanding of stellar evolution have so accelerated the rate of progress, that "nucleogenesis" has now become a well-defined branch of astrophysics. Since its techniques are not directly cosmological, we shall not discuss this work here. Fortunately, some very up to date survey articles have appeared, to which we refer the reader. The observed abundance distribution is discussed by Suess and Urey (1956), Ahrens (1957), Minnaert (1957), Aller (1959). Nucleogenesis itself is surveyed by E. M. Burbidge, G. R. Burbidge, Fowler and Hoyle (1956, 1957), Cameron (1957a, b, c), Greenstein (1956), Fowler and Greenstein (1956). See also: Burbidge and Burbidge (1957); Cook, Fowler, Lauritsen and Lauritsen (1957); Heller (1957); Salpeter (1957) and Aten (1957) for recent papers on this subject.

We should perhaps mention one remarkable result which strongly suggests that nucleogenesis is developing along the right lines. This is Hoyle's (1954) prediction that the C^{12} nucleus must have an excited state at 7·70 MeV in order that the relative abundance of He^4, C^{12}, and O^{16} be accounted for. At the time Hoyle made this prediction there was conflicting evidence about a C^{12} level in the vicinity of 7·7 MeV. Possible values ranged as low as 7·3 MeV (Guggenheimer, Heitler and Powell 1947), while some investigators failed to find the level at all. After Hoyle's prediction, the matter was reinvestigated and the level was indeed found—the observed energy being $7{\cdot}68 \pm 0{\cdot}03$ MeV (Hoyle, Dunbar, Wenzel and Whaling, 1953), $7{\cdot}66 \pm 0{\cdot}02$ (Pauli, 1955), $7{\cdot}658 \pm 0{\cdot}027$ Ahnlund (1956).

It should be emphasized that if this approach is successful, so that all the details of the abundance curve and its variation from place to place could be accounted for by element synthesis in stars, the evolving models of the Universe would not be

disproved, since one can presumably choose initial conditions in these models for which a negligible amount of element-formation occurs in their early stages. On the other hand, if stellar synthesis is an inadequate source of heavy elements, the steady-state model will be ruled out. It will be interesting to see whether the steady-state model survives this test.

(vi) *Prospects for the Future*

It is always dangerous to prophesy, but it does seem as though the most potent instrument for cosmological research may well be radio astronomy; partly because radio methods can detect objects too distant to be detected optically, and partly just because we know so little of its full scope. This is particularly true of 21 cm studies, which should soon lead to the measurement of clusters whose red-shift z exceeds 0·5.

Another potentially powerful technique is the use of artificial satellites, both directly in detecting the light of the night sky (Baum, 1956), and indirectly in extending and making more reliable the whole range of astrophysical data.

Despite the great difficulties, the next decade promises to be an exciting one for observational cosmology—one in which the correct model of the Universe may at last be determined.

VOL. 3 196

REFERENCES

Books:

Aller, L. H. . . . . 1959 *Chemical Composition of the Universe*. Interscience Press, New York; (in press).

Bondi, H. . . . . 1952 *Cosmology* (Cambridge).

Eddington, A. S. . . . . 1926 *The Internal Constitution of the Stars* (Cambridge).

Heckmann, O. . . . . 1942 *Theorien der Kosmologie* (Berlin).

Hubble, E. P. . . . . 1936 *The Realm of the Nebulae* (Oxford).

Jeans, J. . . . . 1928 *Astronomy and Cosmogony* (Cambridge).

McVittie, G. C. . . . . 1956 *General Relativity and Cosmology* (New York).

Schrödinger, E. . . . . 1956 *Expanding Universes* (Cambridge).

Sciama, D. W. . . . . 1959 *The Unity of the Universe*, Faber and Faber, London.

Solvay Conference . . . . 1958 *Institut International de Physique Solvay;* 11th Conseil de Physique: "La Structure et l'Evolution de l'Univers"; 309 pp; Brussels, June 1958.

Whitrow, G. J. . . . . 1959 *The Structure and Evolution of the Universe*, Hutchinson, London.

Zwicky, F. . . . . 1957 *Morphological Astronomy* (Berlin).

(Université de Liège) . . . . 1954 *Les Processus Nucléaires dans les Astres* (Louvain).

Papers:

Abell, G. O. . . . . 1957 *Astron. J.*, **62**, 2.

1958 *Ap. J. Suppl.*, **3**, 211.

Ahnlund, K. . . . . 1956 *Ark. Fys.*, **10**, 369.

Ahrens, L. H. . . . . 1957 in *Physics and Chemistry of the Earth*, Vol. 2, p. 30; (Pergamon Press, London).

Aller, L. H. . . . . 1958 in *Handbuch der Physik* (Springer, Göttingen), Vol. 51, pp. 324–352.

Alpher, R. A., Bethe, H. A. and Gamow, G. 1948 *Phys. Rev.*, **73**, 803.

ALPHER, R. A. and HERMAN, R. C.	1950	*Rev. mod. Phys.*, **22**, 153.
	1953	*Ann. Rev. Nucl. Sci.*, **2**, 1.
		Physica **23**, 1073.
BAADE, W. and SPITZER, L.	1951	*Ap. J.* **113**, 413.
BAADE, W.	1956	*Publ. Astron. Soc. Pacific*, **68**, 5.
BAUM, W. A.	1956	*Publ. Astron. Soc. Pacific*, **68**, 118.
	1957	*Astron. J.*, **62**, 6.
BOLTON, J. G., STANLEY, G. J. and SLEE, O. B.	1954	*Austral. J. Phys.*, **7**, 110.
BOLTON, J. G.	1956a	*Observatory*, **76**, 62.
	1956b	*Publ. Astron. Soc. Pacific*, **68**, 477.
BONDI, H. and GOLD, T.	1948	*M.N.*, **108**, 252.
BONDI, H., GOLD, T. and SCIAMA, D. W.	1954	*Ap. J.*, **120**, 597.
BONNOR, W. B.	1957a	*M.N.*, **117**, 104.
	1957b	*Ann. Inst. Henri Poincaré*, **15**, 158.
BURBIDGE, E. M. and BURBIDGE, G. R.	1957	*Ap. J.*, **126**, 357.
BURBIDGE, E. M., BURBIDGE, G. R., FOWLER, W. A. and HOYLE, F.	1956	*Science*, **124**, 611.
	1957	*Rev. mod. Phys.*, **29**, 547.
CAMERON, A. G. W.	1957a	*Publ. Astron. Soc. Pacific*, **69**, 201.
	1957b	*Astron. J.*, **62**, 6, 138.
	1957c	AECL Report No. 454 (Chalk River), Ontario.
COOK, C. W., FOWLER, W. A. and LAURITSEN, C. C. and T.	1957	*Phys. Rev.*, **107**, 508.
DIETER, N. H.	1957	Doctoral Thesis, Harvard University (March 1958); see also L. VOLDERS and H. C. VAN DE HULST in Paper 78, p. 426, of the IAU 1958 "Paris Symposium on Radio-Astronomy", ed. R. N. BRACEWELL Stanford Univ. Press, 1959.
EDGE, D. O., SCHEUER, P. A. G. and SHAKESHAFT, J. R.	1958	*M.N.* **118**, 183.
EGGEN, O. J., GASCOIGNE, S. C. B. and BURR, E. J.	1957	*M.N.*, **117**, 406.
FIELD, G. B.	1958	*Proc. Inst. Radio Engrs.*, **46**, 240.
	1959	*Ap. J.*, **129**, 525-564 (3 papers).
FARKAS, L. and HARTECK, P.	1931	*Naturwissenschaften*, **19**, 705.
FOWLER, W. A., and GREENSTEIN, J. L.	1956	*Proc. Nat. Acad. Sci.* U.S.A., **42**, 173.
GAMOW, G.	1946	*Phys. Rev.*, **70**, 572.
	1949	*Rev. mod. Phys.* **21**, 367.
	1953	*K. Danske Vidensk Selsk., Mat-fys.*, **27**, No. 10.
GREENSTEIN, J. L.	1954	Liège, 8° Coll., No. 357, p. 307.
	1956	*Publ. Astron. Soc. Pacific.* **68**, 185.
GUGGENHEIMER, K. M., HEITLER, W., and POWELL, C. F.	1947	*Proc. Roy. Soc.*, **190**, 196.
HARO, G.	1956a	*Astron. J.*, **61**, 178.
	1956b	*Bull. Tonanzintla Obs.* No. 14.
HAZARD, C. and WALSH, D.	1958	Paper 87, pp. 477-486 in the IAU 1958 "Paris Symposium on Radio-Astronomy", ed. R. N. Bracewell, Stanford Univ. Press, 1959.
HEESCHEN, D. S.	1956	*Ap. J.*, **124**, 660.
	1957	*Publ. Astron. Pacific*, **69**, 350.
HEESCHEN, D. S. and DIETER, N. H.	1958	*Proc. Inst. Radio Engrs.*, **46**, 234.
HELLER, L.	1957	*Ap. J.*, **126**, 341.
HERZOG, E., WILD, P. and ZWICKY, F.	1957	*Publ. Astron. Soc. Pacific*, **69**, 409.
HOYLE, F.	1948	*M.N.*, **108**, 372.
	1949	*M.N.*, **109**, 365.
	1953	*Ap. J.*, **118**, 513.
HOYLE, F., DUNBAR, D. N. F., WENZEL, W. A. and WHALING, W.	1953	*Phys. Rev.*, **92**, 1095.
HOYLE, F.	1954	*Ap. J., Suppl.* **1**, 121.
HOYLE, F. and SANDAGE, A. R.	1956	*Publ. Astron. Soc. Pacific*, **68**, 301.
HOYLE, F. and HASELGROVE, B.	1956	*M.N.*, **116**, 527.
HUMASON, M. L., MAYALL, N. U. and SANDAGE, A. R.	1956	*Astron. J.*, **61**, 97.
JUST, R.	1959	*Ap. J.*, **129**, 268.
LAYZER, D.	1954	*Astron. J.*, **59**, 170.
	1956	*Astron. J.*, **61**, 383 (and unpublished work).
LEMAÎTRE, G.	1949	*Rev. mod. Phys.*, **21**, 357.
LILLEY, A. E. and MCCLAIN, E. F.	1956	*Ap. J.*, **123**, 172.
	1958	*Proc. Inst. Radio Engrs.*, **46**, 221.
LIMBER, D. N.	1957	*Ap. J.* **125**, 9.
MCCREA, W. H.	1950	*Endeavour*, **9**, 3.

McCrea, W. H.	1951	*Proc. Roy. Soc.*, **206**, 562.
	1953	*Rep. Progr. Phys.*, **16**, 321.
	1958	*Endeavour*, **17**, 5.
McClain, E. F.	1956	*Ap. J.*, **123**, 367.
Merrill, P. W.	1952	*Science*, **115**, 484.
Mestel L. and Spitzer, L.	1956	*M.N.*, **116**, 503.
Mills, B. Y. and Little, C. G.	1953	*Austral. J. Phys.*, **6**, 272.
Mills, B. Y. and Slee, O. B.	1957	*Austral. J. Phys.*, **10**, 162.
Minkowski, R. and Wilson, O. C.	1956	*Ap. J.*, **123**, 373.
Minkowski, R.	1958	*Publ. Astron. Soc. Pacific*, **70**, 143.
Minnaert, M. G. J.	1957	*M.N.*, **117**, 315.
Morgan, W. W. and Mayall, N. U.	1957	*Publ. Astron. Soc. Pacific*, **69**, 291.
Münch, G.	1957	Director's Report, Mount Wilson and Palomar Observatories.
	1958	*Sky and Telescope*, **17**, 231.
Neyman, J.	1955	*Ann. Inst. Henri Poincaré*, **14**, 201.
Neyman, J., Scott, E. L. and Shane, C. D.	1956	*Proc. Third Berkeley Symposium Math. Statistical Problems*, p. 75.
Neyman, J. and Scott, E. L.	1959a	in *Handbuch der Physik*. Vol. 53, pp. 416–444.
	1959b	*J. Roy. Statistical Soc.* To be published.
Pauli, R. T.	1955	*Ark. Fys.*, **9**, 571.
Pawsey, J. L.	1958	*Publ. Astron. Soc. Pacific*, **70**, 133.
Pokrowski, G. I.,	1931	*Phys., Z.* **32**, 374.
Purcell, E. M. and Field, G. B.	1956	*Ap. J.*, **124**, 542.
Robertson, H. P.	1955	*Publ. Astron. Soc., Pacific*, **67**, 82.
	1956	*Helv. phys. Acta., Suppl.* 4.
Ryle, M. and Scheuer, P. A. G.	1955	*Proc. Roy. Soc.*, **230**, 448.
Ryle, M. and Hewish, A.	1955	*Mem. Roy. Astron. Soc.*, **67**, 97.
Ryle, M.	1958	*Proc. Roy. Soc. A* **248**, 289.
Salpeter, E. E.	1957	*Phys. Rev.*, **107**, 516.
Sandage, A. R.	1957	*Ap. J.*, **126**, 326.
	1958	*Sky and Telescope*, **17**, 275.
Sané, M.	1957	*A.J.*, **62**, 348.
Scheuer, P. A. G.	1957	*Proc. Cambridge Phil. Soc.*, **53**, 764.
Schwarzschild, M. and B., Searle, L. and Meltzer, A.	1957	*Ap. J.*, **125**, 123.
Sciama, D. W.	1955	*M.N.*, **115**, 3.
Scott, E. L.	1955	*Astron. J.*, **60**, 292.
	1957	*Astron. J.*, **62**, 248.
Shakeshaft, J. R., Ryle, M. Baldwin, J. E., Elsmore B. and Thomson, J. H.	1955	*Mem. Roy. Astron. Soc.*, **67**, 97.
Shane, C. D.	1955	*Astron. J.*, **60**, 292; another paper to be published.
	1956	in *Vistas in Astronomy*, Ed. A. Beer, Vol. 2, p. 1574; (Pergamon Press, London).
Stebbins, J. L. and Whitford, A. E.	1948	*Ap. J.*, **108**, 413.
Stock, J. and Schuecking, E.	1957	*Astron. J.*, **62**, 98.
Suess, H. E. and Urey, H. C.	1956	*Rev. mod. Phys.*, **28**, 53.
Vaucouleurs, G. de	1948	*C. R. Acad. Sci.* (Paris), **227**, 466.
	1956	in *Vistas in Astronomy*, Ed. A. Beer, Vol. 2, p. 1574; (Pergamon Press, London).
	1958	*Sky and Telescope*, **17**, 232.
Walker, A. G.	1944	*J. Lond. Math. Soc.*, **19**, 219, 227.
Whitford, A. E.	1954a	*Ap. J.*, **120**, 599.
	1954b	*Astron. J.*, **59**, 194.
	1956	*Astron. J.*, **61**, 352.
Whitford, A. E. and Code, A. D.	1957	*Sky and Telescope*, **16**, 222.
Whitrow, G. J.	1954	*M.N.*, **114**, 180.
Zwicky, F.	1956	*Publ. Astron. Soc. Pacific*, **68**, 331.
	1957	*Publ. Astron. Soc. Pacific*, **69**, 518'
	1959	in *Handbuch der Physik*, (Springer, Gottingen), Vol. 53, pp. 373–415.

VOL
3
196

CONTENTS OF PREVIOUS VOLUMES

CONTENTS OF VOLUME 1

Introduction *by* ARTHUR BEER

Hommage respectueux, cordial, et reconnaissant à Monsieur le Professeur F. J. M. STRATTON: PLANET 1942 XB = "STRATTONIA" *by* E. DELPORTE,† (Bruxelles)

1. CO-OPERATION AND ORGANIZATION

International co-operation in astronomy *by* M. MINNAERT (Utrecht)
The International Astronomical Union *by* P. TH. OOSTERHOFF (Leiden)
The International Astronomical News Service *by* JULIE VINTER HANSEN (Copenhagen)
The Committee for the Distribution of Astronomical Literature and the Astronomical News Letters *by* BART J. BOK (Canberra) and V. KOURGANOFF (Lille)
Some problems of international co-operation in geophysics *by* J. M. STAGG (London)
Some educational aspects of astronomy *by* V. BAROCAS (Preston, Eng.)
Recollections of seventy years of scientific work *by* J. EVERSHED, F.R.S.,† (Ewhurst, Eng.)
Astronomical recollections *by* MAX BORN, F.R.S. (Bad Pyrmont, Germany)

2. HISTORY AND PHILOSOPHY

The Egyptian "Decans" *by* O. NEUGEBAUER (Providence, R.I.)
The astral religion of antiquity and the "Thinking Machines" of today *by* GEORGE SARTON,† (Cambridge, Mass).
Ptolemy's precession *by* A. PANNEKOEK (Wageningen, Netherlands)
A medieval footnote to Ptolomaic precession *by* DEREK J. PRICE (New Haven, Conn.)
The Peking Observatory in A.D. 1280 and the development of the equatorial mounting *by* JOSEPH NEEDHAM, F.R.S. (Cambridge)
The Mercury Horoscope of Marcantonio Michiel of Venice: a study in the history of Renaissance astrology and astronomy *by* WILLY HARTNER (Frankfurt a.M.)
Observatories and instrument makers in the eighteenth century *by* D. W. DEWHIRST (Cambridge)
The Radcliffe Observatory *by* H. KNOX-SHAW (Elgin, South Africa)
A system of quantitative astronomical observations *by* C. W. ALLEN (London)
Fact and inference in theory and in observation *by* H. BONDI, F.R.S. (London)
Philosophical aspects of cosmology *by* HERBERT DINGLE (London)
Modern cosmology and the theologians *by* M. DAVIDSON (Audley End, Eng.)

3. DYNAMICS

Vistas in celestial mechanics *by* D. H. SADLER (Herstmonceux)
An introduction to the eclipse Moon *by* R. d'E. ATKINSON (Herstmonceux)
The Moon's principal librations in rectangular co-ordinates *by* SIR HAROLD JEFFREYS, F.R.S. (Cambridge)
A general expression for a Lagrangian Bracket *by* W. M. SMART (Glasgow)
The rôle of Cracovians in astronomy *by* T. BANACHIEWICZ,† (Cracow)
Regularization of the Three-Body Problem *by* G. LEMAÎTRE (Bruxelles)
Recent developments in stellar dynamics *by* G. L. CAMM (Manchester)
"Intrinsic" studies of stellar movements in the Milky Way *by* PAUL BOURGEOIS (Bruxelles)
The *K*-effect in stellar motions *by* HAROLD F. WEAVER (Berkeley, Cal.)
On the empirical foundation of the General Theory of Relativity *by* E. FINLAY-FREUNDLICH (Wiesbaden)

4. THEORETICAL ASTROPHYSICS

Fraunhofer lines and the structure of stellar atmospheres *by* A. Unsöld (Kiel) and V. Weidemann (Brunswick)
On the formation of condensations in a gaseous nebula *by* H. Zanstra (Amsterdam)
The calculation of atomic transition probabilities *by* R. H. Garstang (London)
Atomic collision processes in astrophysics *by* H. S. W. Massey, F.R.S. (London)
Energy production in stars *by* E. E. Salpeter (Ithaca, N.Y.)
The effective pressure exerted by a gas in turbulent motion *by* G. K. Batchelor (Cambridge)
"Turbulence," kinetic temperature, and electron temperature in stellar atmospheres *by* P. L. Bhatnagar (Delhi), M. Krook (Cambridge, Mass.), D. H. Menzel (Cambridge, Mass.) and R. N. Thomas (Boulder, Col.)
The intensity of emission lines in stellar spectra *by* P. Wellmann (Hamburg-Bergedorf)
Dynamo theories of cosmic magnetic fields *by* T. G. Cowling, F.R.S. (Leeds)
On the interpretation of stellar magnetic fields *by* S. K. Runcorn (Newcastle-upon-Tyne)
Theories of variable stellar magnetic fields *by* V. C. A. Ferraro (London)
Theories of interstellar polarization *by* Leverett Davis, Jr. (Pasadena, Cal.)
The gravitational instability of an infinite homogeneous medium when a Coriolis acceleration is acting *by* S. Chandrasekhar, F.R.S. (Williams Bay, Wis.)

5. INSTRUMENTS

Modern developments in telescope optics *by* E. H. Linfoot (Cambridge)
Thermal distortions of telescope mirrors and their correction *by* André Couder (Paris)
On the interferometric measurement of small angular distances *by* André Danjon (Paris)
Partially coherent optical fields *by* E. Wolf (Rochester, N.Y.)
On a possible improvement of Michelson's method for the determination of stellar diameters in poor visibility *by* R. Fürth (London) and E. Finlay-Freundlich (Wiesbaden)
Astronomical spectrographs: past, present and future *by* I. S. Bowen (Pasadena, Cal.)
Spectroscopy with the echelle *by* George R. Harrison (Cambridge, Mass.)
Comparison of the efficiency of prism- and grating-spectrographs *by* Ch. Fehrenbach (Marseille)
Optical systems employed in spectroheliographs *by* P. A. Wayman (Cape)
The solar installation of Dunsink Observatory *by* H. A. and Mary T. Brück (Edinburgh)
Quartz crystal clocks *by* Humphrey M. Smith and G. B. Wellgate (Herstmonceux)
A rapid method of obtaining positions from wide-angle plates *by* G. van Biesbroeck (Williams Bay, Wis.)
Nomograms in astronomy *by* H. Strassl (Münster)
The use of electronic calculating machines in astronomy *by* Peter Naur (Copenhagen)
Air disturbance in reflectors *by* W. H. Steavenson (Cambridge)
Photoelectric devices in astronomy *by* P. B. Fellgett (Edinburgh)
The development of photoelectric photometry *by* C. M. Huffer (Madison, Wis.)
On the application of the photomultiplier to astronomical photometry *by* R. H. Hardie (Nashville, Tenn.)
On ideal reception of photons and the electron telescope *by* A. Lallemand (Paris)
Indirect methods of star counting *by* H. E. Butler (Edinburgh)

6. RADIO ASTRONOMY

Solar radio eclipse observations *by* J. S. Hey (Malvern, Eng.)
The application of interferometric methods in radio astronomy *by* M. Ryle, F.R.S. (Cambridge)
Large radio telescopes and their use in radio astronomy *by* R. Hanbury Brown and A. C. B. Lovell, F.R.S. (Manchester)
Positional radio-astronomy and the identification of some of the radio stars *by* F. G. Smith (Cambridge)
Australian work on radio stars *by* J. G. Bolton (Pasadena)
Outbursts of radio noise from the Sun *by* J. P. Wild (Sydney)
Radio echo studies of meteors *by* J. G. Davies and A. C. B. Lovell, F.R.S. (Manchester)
The scintillation of radio stars *by* A. Hewish (Cambridge)
Measures of the 21-cm line emitted by interstellar hydrogen *by* J. H. Oort (Leiden)

7. SOLAR PHYSICS

Early solar research at Mount Wilson *by* W. S. Adams, For.Mem.R.S.† (Pasadena, Cal.)
Thirty years of solar work at Arcetri *by* Giorgio Abetti (Arcetri-Firenze)
Some problems of the international solar survey *by* K. O. Kiepenheuer (Freiburg i.B.)
Physical conditions in the solar photosphere *by* H. H. Plaskett, F.R.S. (Oxford)
Infra-red Fraunhofer lines and photospheric structure *by* B. E. J. Pagel (Herstmonceux)
Temperature at the poles and at the equator of the Sun *by* A. K. Das (Kodaikanal, India) and K. D. Abhyankar (Berkeley, Cal.)
The lineage of the great sunspots *by* H. W. Newton (Herstmonceux)
The structure of sunspots *by* P. A. Sweet (Glasgow)
The magnetic fields of sunspots *by* Sir Edward C. Bullard, F.R.S. (Cambridge)
On mean latitudinal movement of sunspots *by* Jaakko Tuominen (Helsinki)
The evolution of solar prominences *by* L. d'Azambuja (Meudon)
Investigations of flares and prominences at the Crimean Astrophysical Observatory *by* A. B. Severny (Simeis, Crimea)
Eclipse observing *by* Sir John A. Carroll (London)
Recent eclipse work in Japan *by* Yusuke Hagihara (Tokyo)
The widths of narrow lines in the spectrum of the low chromosphere, measured at the total eclipse of 25 February 1952 *by* R. O. Redman, F.R.S. (Cambridge)
The excitation temperatute of the solar chromosphere determined from molecular spectra *by* D. E. Blackwell (Cambridge)
Solar work and plans in Egypt *by* M. R. Madwar (Cairo)
Spectroscopic studies of the solar corona *by* G. Righini (Arcetri-Firenze)
Coronagraphic work at moderate altitude above sea-level *by* Yngve Öhman (Stockholm)
Spectroscopic measurements of magnetic fields on the Sun *by* H. von Klüber (Cambridge)

CONTENTS OF VOLUME 2

8. SOLAR-TERRESTRIAL RELATIONS

Regularities and irregularities in the ionosphere *by* Sir Edward Appleton, F.R.S. (Edinburgh)
The microscopic mechanism for the absorption of radio waves in the ionosphere *by* J. A. Ratcliffe, F.R.S. (Cambridge)
Solar flares *by* M. A. Ellison (Dunsink-Dublin)
M-regions and solar activity *by* M. Waldmeier (Zürich)
The corpuscular radiation of the Sun *by* M. J. Smyth (Edinburgh)
Cosmic rays and the Sun *by* P. M. S. Blackett, F.R.S. (London)

9. GEOPHYSICS

Non-seasonal changes in the Earth's rotation *by* Sir Harold Spencer Jones, F.R.S., late Astronomer Royal (Bexhill-on-Sea, Sussex)
Seasonal fluctuations in the rate of rotation of the Earth *by* F. H. van den Dungen, F. J. Cox and J. van Mieghem (Bruxelles)
The thickness of the continents *by* Robert Stoneley, F.R.S. (Cambridge)
Some techniques for locating thunder-storms from a single observing station *by* E. T. Pierce (Boston, Mass.)
The abundance of atmospheric carbon dioxide and its isotopes *by* Leo Goldberg (Ann Arbor, Mich.)
The infra-red spectrum of the Earth's atmosphere *by* T. W. Wormell (Cambridge)
The ozone layer: an investigation of the 9.6 μ ozone band in the telluric spectrum *by* R. M. Goody (Cambridge, Mass.) and C. D. Walshaw (Cambridge, Eng.)
Atmospheric ozone near the North Pole *by* A. E. Douglas, G. Herzberg, F.R.S. and D. C. Rose (Ottawa)
Research in the upper atmosphere with high altitude sounding rockets *by* S. F. Singer (College Park, Maryland)
The morphology of geomagnetic storms and bays: a general review *by* Sydney Chapman, F.R.S. (College, Alsska)
The airglow *by* Daniel Barbier (Paris)

10. PLANETARY SYSTEM

Current trends in minor planets' research *by* DIRK BROUWER (New Haven, Conn.)
On the present situation in cometary research *by* M. BEYER (Hamburg-Bergedorf)
The spectra of the comets *by* P. SWINGS (Liège)
The Harvard photographic meteor programme *by* LUIGI G. JACCHIA and FRED L. WHIPPLE (Cambridge, Mass.)
On the evolution of meteor streams *by* MIROSLAV PLAVEC (Praha)
The zodiacal light *by* H. C. VAN DE HULST (Leiden)

11. STELLAR ASTRONOMY

Recent advances in positional astronomy *by* A. KOPFF (Heidelberg)
On the increase of the accuracy of fundamental declinations *by* F. SCHMEIDLER (München)
The distances of the stars: a historical review *by* J. JACKSON, F.R.S.,† (Ewell, Eng.)
Fifty years of trigonometrical parallaxes *by* A. HUNTER and E. G. MARTIN (Herstmonceux)
The internal accuracy of the Yale parallaxes *by* GUSTAV LAND (New Haven, Conn.)
The visual binaries *by* W. H. VAN DEN BOS (Johannesburg)
Planetary companions of stars *by* PETER VAN DE KAMP (Swarthmore, Pa.)
White dwarfs and degenerate stars *by* WILLIAM J. LUYTEN (Minneapolis, Minn.)
The dimensions of Omega Centauri *by* E. M. LINDSAY (Armagh, N. Ireland)
On the groups of diffuse emission nebulae *by* G. A. SHAJN,† (Moscow)
The problem of Hagen's Clouds *by* FRIEDRICH BECKER and J. MEURERS (Bonn)
Interstellar dust *by* G. ALTER (Praha)
Interstellar polarization *by* W. A. HILTNER (Williams Bay, Wis.)

12. PHOTOMETRY

Monochromatic magnitudes *by* R.V.D.R. WOOLLEY, F.R.S., Astronomer Royal (Herstmonceux)
The Cape magnitudes and their relationship to the international system *by* R. H. STOY (Cape)
The zero-point in stellar photometry *by* E. RYBKA (Wroclaw)
Photographic photometry of the Hyades *by* O. HECKMANN (Hamburg-Bergedorf)
Note on the diameters of photovisual star images *by* EJNAR HERTZSPRUNG (Tølløse, Denmark)
The galactic cluster M 29 (NGC 6913) *by* W. W. MORGAN and D. L. HARRIS (Williams Bay, Wis.)
A contribution to the problem of the classification of variable stars *by* LIVIO GRATTON (La Plata, Argentine)
Population II Cepheids of the Galaxy *by* CECILIA PAYNE-GAPOSCHKIN (Cambridge, Mass.)
Spectrophotometric observations of Cepheids at Arcetri *by* MARGHARITA HACK (Milano-Merate)
The problem of RR Lyrae stars with several periods *by* L. DETRE (Budapest)
Photographic light-curve of Eta Carinae *by* D. J. K. O'CONNELL, S.J. (Castelgandolfo, Vatican City)
Eclipsing variables—a survey of the field *by* FRANK BRADSHAW WOOD (Philadelphia, Pa.)
Eclipsing variables in the Magellanic Clouds *by* HENRY NORRIS RUSSELL, For.Mem.R.S.† (Princeton, N.J.)
On finding the apsidal period of an eclipsing binary *by* J. DE KORT, S.J. (Castelgandolfo, Vatican City)
Photometric effects of gaseous envelopes in close eclipsing binaries *by* MICHAEL W. OVENDEN (Glasgow)
Statistics and recent results on eclipsing binaries *by* MARIO G. FRACASTORO (Catania)

13. SPECTROSCOPY

Atomic spectra—their rôle in astrophysics *by* CHARLOTTE E. MOORE (Washington, D.C.)
Methods in stellar spectroscopy *by* THEODORE DUNHAM, Jr. (Canberra)
The quantitative chemical analysis of early-type stars *by* LAWRENCE H. ALLER (Ann Arbor, Mich.)
The technique and possibilities of astronomical spectrophotometry *by* JESSE L. GREENSTEIN (Pasadena, Cal.)
The continuous spectrum *by* W. M. H. GREAVES, F.R.S.† (Edinburgh)
The problem of stellar temperatures *by* H. KIENLE (Heidelberg)
New spectral classifications with two or three parameters *by* D. CHALONGE (Paris)

VOL.
3
196(

Two-dimensional spectral classification of F stars through photoelectric photometry with interference filters *by* BENGT STRÖMGREN (Princeton, N.J.)
Luminosities of the B-stars from spectroscopic measurements *by* R. M. PETRIE (Victoria, B.C.)
Luminosity classification of stars of late spectral type from spectra of moderate dispersion *by* JOHN F. HEARD (Toronto)
Objective-prism spectra of relatively cool stars in the near infrared *by* J. J. NASSAU (Cleveland, Ohio)

14. SPECTRAL PECULIARITIES AND NOVAE

Spectroscopic phenomena in the system of Algol *by* OTTO STRUVE, For.Mem.R.S. (Green Bank, West Virginia)
Stars with expanding atmospheres *by* PAUL W. MERRILL (Pasadena, Cal.)
Shells around P Cygni stars *by* A. D. THACKERAY (Pretoria)
The masses of spectroscopic binaries *by* ARTHUR BEER (Cambridge)
Evidence for and against binary hypotheses of the blue-red erupting stars *by* MARTIN JOHNSON (Birmingham)
Observational results on variable stars with composite spectra *by* TCHENG MAO-LIN (Peking) and M. BLOCH (Lyon)
The peculiar A-stars *by* ARMIN J. DEUTSCH (Pasadena, Cal.)
The carbon stars—an astronomical enigma *by* WILLIAM P. BIDELMAN (Mount Hamilton, Cal.)
Celestial emission line objects of early type *by* C. S. BEALS, F.R.S., and J. A. ROTTENBERG (Ottawa)
Emission lines and stellar atmospheres *by* G. R. BURBIDGE and E. M. BURBIDGE (Pasadena)
Identification of emission lines in the spectra of Wolf-Rayet stars *by* BENGT EDLÉN (Lund)
Gamma Cassiopeiae *by* D. L. EDWARDS,† (Sidmouth, Eng.)
The spectra of novae *by* DEAN B. MCLAUGHLIN (Ann Arbor, Mich.)
The mechanism of ejection of matter from novae *by* E. R. MUSTEL (Simeis, Crimea)
On the nature of physical processes underlying the nova phenomenon *by* ZDENĚK KOPAL (Manchester)
Novae observed *by* SERGEI GAPOSCHKIN (Cambridge, Mass.)
On the relics of super-novae *by* W. H. RAMSEY (Manchester)

15. GALAXIES

An analysis of the Milky Way *by* WILHELM BECKER (Basel)
Milky Way structure in Sagittarius and Carina; a study in contrasts *by* BART J. BOK (Canberra)
Infra-red photography in the Milky Way, particularly in Sagittarius *by* JEAN DUFAY (Lyon), JOSEPH-HENRI BIGAY (Lyon), and PIERRE BERTHIER (St. Michel, B.-Alpes)
The great nebulae of the Southern sky *by* DAVID S. EVANS (Cape)
Galactic and extragalactic studies at the Harvard Observatory *by* HARLOW SHAPLEY (Cambridge, Mass.)
The distribution of the galaxies *by* C. D. SHANE (Mount Hamilton, Cal.)
The distribution of bright galaxies and the Local Supergalaxy *by* GÉRARD DE VAUCOULEURS (Cambridge, Mass.)
On metagalactic distance-indicators *by* KNUT LUNDMARK,† (LUND)
Red-shifts in the spectra of extragalactic nebulae by MILTON L. HUMASON (Pasadena, Cal.)

16. COSMOGONY AND COSMOLOGY

The origin of the satellites and the Trojans *by* GERARD P. KUIPER (Williams Bay, Wis.)
The origin and significance of the Moon's surface *by* HAROLD C. UREY, For.Mem.R.S. (Chicago, Ill.)
The frequency of meteorite falls throughout the ages *by* F. A. PANETH, F.R.S.,† (Mainz)
The essential nature of accretion by R. A. LYTTLETON, F.R.S. (Cambridge)
On the production of groups of O- and B-stars by accretion *by* W. H. MCCREA, F.R.S. (London)
The effect of galactic rotation on accretion by main sequence stars *by* F. HOYLE, F.R.S. (Cambridge)
On the origin of double stars *by* V. A. AMBARTSUMIAN (Bjurakan, Armenia)
On the evolution of stellar systems *by* BERTIL LINDBLAD (Stockholm)
Cosmology *by* T. GOLD (Ithaca, N.Y.)
The physics of the expanding Universe *by* G. GAMOW (Boulder, Col.)

NAME INDEX
SUBJECT INDEX